The American Economic Review

CYK-H03126-69-P000954 69

ARTICLES

W. W. HELLER	What's Right With Economics?
R. B. FREEMAN	Supply and Salary Adjustments to the Changing Science Manpower Market: Physics, 1948–1973
B. B. AGHEVLI	The Balance of Payments and the Money Supply Under the Gold Standard Regime: U.S. 1879–1914
R. M. COEN	Investment Behavior, The Measurement of Depreciation, and Tax Policy
M. S. FELDSTEIN	Wealth Neutrality and Local Choice in Public Education
W. L. GRAMM	Household Utility Maximization and the Working Wife
O. H. SCHENONE	A Dynamic Analysis of Taxation
D. A. GRAHAM	A Geometrical Exposition of Input-Output Analysis
L. DE ALESSI	Toward an Analysis of Postdisaster Cooperation
J. M. BUCHANAN AND G. TULLOCK	Polluters' Profits and Political Response: Direct Controls Versus Taxes
T. T. SU	Durability of Consumption Goods Reconsidered
J. G. HAWORTH, J. GWARTNEY, AND C. HAWORTH	Earnings, Productivity, and Changes in Employment Discrimination During the 1960's
G. DE MENIL AND S. S. BHALLA	Direct Measurement of Popular Price Expectations
G. A. CALVO	Efficient and Optimal Utilization of Capital Services

CALCUTTA UNIVERSITY

SHORTER PAPERS: R. E. Baldwin; R. Hendler; E. James; R. Eisner and P. J. Lawler; N. S. Barrett; R. J. Flanagan; L-S. Fan; D. Orr; T. Horst; C. R. Link; L. Weiss and J. G. Williamson; R. M. Korkie; H. E. Leland; M. C. Kemp and E. Tower; R. N. Batra.

MARCH 1975

THE AMERICAN ECONOMIC ASSOCIATION

Founded in 1885

- Published at George Banta Co., Inc., Menasha, Wisconsin.

- The American Economic Review, including four quarterly numbers, the *Proceedings* of the annual meetings, the *Directory*, and Supplements, is published by the American Economic Association and is sent to all members five times a year, in March, May, June, September, and December.

- Membership dues of the Association are $23.00 a year, which includes a year's subscription to both the *American Economic Review* and the *Journal of Economic Literature*. Junior membership at $11.50 per year is available to registered students for a maximum of three years. Subscriptions (libraries, institutions, or firms) are $34.50 a year, and only subscriptions to both publications will be accepted. Please write the Secretary's office, Nashville, Tennessee for single copy prices.

- Correspondence relating to the *Papers and Proceedings*, the *Directory*, advertising, permission to quote, business matters, subscriptions, membership and changes of address should be sent to the Secretary, Rendigs Fels, 1313 21st Avenue South, Nashville, Tennessee 37212. Change of address must reach the Secretary at least six (6) weeks prior to the month of publication. The Association's publications are mailed second class and are not forwarded by the Post Office.

- Second-class postage paid at Nashville, Tennessee and at additional mailing offices. Printed in U.S.A.

Officers

President

ROBERT AARON GORDON
University of California, Berkeley

President-Elect

FRANCO MODIGLIANI
Massachusetts Institute of Technology

Vice-Presidents

MARTIN BRONFENBRENNER
Duke University
CHARLES L. SCHULTZE
University of Maryland

Secretary-Treasurer and Editor of the Proceedings

RENDIGS FELS
Vanderbilt University

Assistant Secretary-Treasurer

C. ELTON HINSHAW
Vanderbilt University

Managing Editor of The American Economic Review

GEORGE H. BORTS
Brown University

Managing Editor of The Journal of Economic Literature

MARK PERLMAN
University of Pittsburgh

Executive Committee

Elected Members of the Executive Committee

IRMA ADELMAN
University of Maryland
ROBERT HEILBRONER
New School for Social Research
ANDREW F. BRIMMER
Harvard University
PAUL M. SWEEZY
Monthly Review
CAROLYN SHAW BELL
Wellesley College
BURTON A. WEISBROD
University of Wisconsin, Madison

Ex Officio Members

KENNETH J. ARROW
Harvard University
WALTER W. HELLER
University of Minnesota

THE AMERICAN ECONOMIC REVIEW

391
- 001

March 1975

VOLUME LXV, NUMBER 1

GEORGE H. BORTS
Managing Editor

WILMA ST. JOHN
Assistant Editor

Board of Editors

BELA BALASSA
EUGENE F. FAMA
MARTIN S. FELDSTEIN
ANN F. FRIEDLAENDER
STEPHEN M. GOLDFELD
ROBERT J. GORDON
ROBERT E. HALL
BENT HANSEN
JAMES R. MELVIN
WILLIAM D. NORDHAUS
LEONARD A. RAPPING
STEPHEN RESNICK
ANNA J. SCHWARTZ
JEROME STEIN
JOSEPH E. STIGLITZ
S. C. TSIANG
FINIS WELCH
MARINA V. N. WHITMAN

- Manuscripts and editorial correspondence relating to the regular quarterly issue of this REVIEW should be addressed to George H. Borts, Managing Editor of THE AMERICAN ECONOMIC REVIEW, Brown University, Providence, R.I. 02912. Manuscripts should be submitted in duplicate and in acceptable form and should be no longer than 50 pages of double-spaced typescript. A $15 submission fee must accompany each manuscript. *Style Instructions* for guidance in preparing manuscripts will be provided upon request to the editor.

- No responsibility for the views expressed by authors in this REVIEW is assumed by the editors or the publishers, The American Economic Association.

- Copyright American Economic Association 1975.

Articles

What's Right With Economics? *Walter W. Heller* 1

Supply and Salary Adjustments to the Changing Science Manpower Market: Physics, 1948–1973 *R. B. Freeman* 27

The Balance of Payments and Money Supply Under the Gold Standard Regime: U.S. 1879–1914 *Bijan B. Aghevli* 40

Investment Behavior, The Measurement of Depreciation, and Tax Policy *Robert M. Coen* 59

Wealth Neutrality and Local Choice in Public Education *Martin S. Feldstein* 75

Household Utility Maximization and the Working Wife *Wendy Lee Gramm* 90

A Dynamic Analysis of Taxation *Osvaldo H. Schenone* 101

A Geometrical Exposition of Input-Output Analysis *Daniel A. Graham* 115

Toward an Analysis of Postdisaster Cooperation *Louis De Alessi* 127

Polluters' Profits and Political Response: Direct Controls Versus Taxes *James M. Buchanan and Gordon Tullock* 139

Durability of Consumption Goods Reconsidered *Teddy T. Su* 148

Earnings, Productivity, and Changes in Employment Discrimination During the 1960's *Joan Gustafson Haworth, James Gwartney, and Charles Haworth* 158

Direct Measurement of Popular Price Expectations *George de Menil and Surjit S. Bhalla* 169

Efficient and Optimal Utilization of Capital Services *Guillermo A. Calvo* 181

Shorter Papers

International Trade and Economic Growth: A Diagrammatic Analysis *Robert E. Baldwin* 187

Lancaster's New Approach to Consumer Demand and Its Limitations *Reuven Hendler* 194

A Note on Uncertainty and the Evaluation of Public Investment Decisions *Estelle James* 200

Tax Policy and Investment: An Analysis of Survey Responses *Robert Eisner and Patrick J. Lawler* 206

The U.S. Phillips Curve and International Unemployment Rate Differentials:
- Comment *Nancy Smith Barrett* 213
- Reply *Robert J. Flanagan* 222

On the Reward System *Liang-Shing Fan* 226

The Industrial Composition of U.S. Exports and Subsidiary Sales to the Canadian Market:
- Comment *Dale Orr* 230
- Reply *Thomas Horst* 235

Black Education, Earnings, and Interregional Migration:
- A Comment and Some New Evidence *Charles R. Link* 236
- Even Newer Evidence *Leonard Weiss and Jeffrey G. Williamson* 241

Theory of the Firm Facing Uncertain Demand:
- Comment *Robert M. Korkie* 245
- Reply *Hayne E. Leland* 248

Nontraded Goods, Factor Market Distortions, and the Gains from Trade:
- Comment *Murray C. Kemp and Edward Tower* 249
- Reply *Raveendra N. Batra* 251

Notes 253

P954

391
·001

Number 76 of a series of photographs of past presidents of the Association

CALCUTTA UNIVERSITY

Walter W. Muller

What's Right With Economics?

By Walter W. Heller*

I come here with no eye-opening report from the frontiers of economics, no stirring cry for reform of conventional economics, no closely reasoned analysis of an economic dilemma or puzzle, no scathing or reproachful scolding of the profession for its technical preciousness or moral blindness, no report on painstaking research results, no valedictory on a lifetime of theoretical or empirical contributions. The AEA presidential addresses have been all of these things.

But tonight, going against our current fashion of telling the world what's wrong with economics, I offer a modest contribution to the immodest subject of what's right with economics—and, in particular, what's right with economics as a guide to public policy. In doing so, I won't ignore the dark side of the moon—indeed, I can't, since I will deal at some length with the bedeviling subject of inflation. But believing that it is at least as reasonable to judge a discipline by its successes as by its failures, I intend to accentuate the positive.

I. The Critical Look Inward

In recent years, as I shall illustrate in a moment, we have instead accentuated the negative. In good part, this has taken the becoming form of *mea culpa* or rather *nostra culpa*. We have, for example, readily confessed that the inflationary shocks of 1973–74 caught not just the economy but the economist by surprise. On this and other fronts, the chorus of self-criticism has risen to a new crescendo. It is almost as if we take pride in our humility. Nietzsche must have been thinking of economists when he observed that "he who despises himself nevertheless esteems himself as a self-despiser."

This is not to imply that economists' criticisms are all self-inflicted wounds. Far from it. Often among our colleagues' favorite targets are the shortcomings of mainstream economics, the misuse of modern techniques, the fallacies of conventional wisdom—in each case, the target is not the critic's but his colleagues' brand of economics, not *mea culpa* but *eorum culpa*.

In any event, he who comes to praise economics risks being buried in the barrage of indictments that economists have brought against themselves and their brethren. Let me give you a sampling of some that will be ringing in my ears as I follow the parlous path of economic virtue.

Ceremonial occasions—presidential, memorial, or inaugural addresses—in particular seem to evoke musings on the troubled or even dismal state of our science. For the AEA faithful, I need only recall John Kenneth Galbraith condemning neoclassical and neo-Keynesian economics for ignoring power and thus losing contact with the real world; Wassily Leontief attacking mathematical economics for building a showy superstructure on weak empirical foundations and unverified assumptions, and thus losing contact with the real world; Kenneth

---

* Presidential address delivered at the eighty-seventh meeting of the American Economic Association, San Francisco, California, December 29, 1974. The address was abridged for oral delivery. I owe particular thanks for the many conversations I held with Francis Boddy, Otto Eckstein, Edward Foster, Arthur Okun, Joseph Pechman, George Perry, Robert Solow, and James Tobin. I also wish to thank Gardner Ackley, Kenneth Arrow, Walter P. Heller, John R. Meyer, Franco Modigliani, Alice Rivlin, Paul Samuelson, Charles Schultze, Christopher Sims, and George Stigler, for contributions to the adult education that underlies this address. The errors of course are mine.

Boulding assailing welfare economics for its reliance on that holiest of holies, Pareto optimality—when in fact "our lives are dominated by precisely this inter-dependence of utility functions which the Paretian optimum denies"—thus also losing contact with the real world.

In one form or another, variations on Leontief's lament have been heard in many another presidential address, to wit:

> By F. H. Hahn (Econometric Society, 1968), who decried "the spectacle of so many people refining the analysis of economic states which they give no reason to suppose will ever, or have ever, come about. . . ."

> By G. D. N. Worswick (Section F of the British Association, 1971), who viewed the performance of economics as "curiously disappointing," suggesting that it has "a marvelous array of pretend tools which would perform wonders if ever a set of facts should turn up in the right form."

> By E. H. Phelps Brown (Royal Economics Society, 1971), who judged the usefulness of current work in economics as "not equal to its distinction" because it is "built upon assumptions about human behavior that are plucked from the air."

> By James H. Blackman (Southern Economic Association, 1971), who noted that models with sufficiently intriguing mathematical properties can achieve lives of their own even if they lead the investigator further away from reality and yet, "the profession's incentive system tends perversely to reward this kind of endeavor and to deflect the attention of gifted economists from the exploration of concrete problems and the dirty work that entails."

> By Sherman Maisel (American Finance Association, 1973), who concluded that most of the literature of monetary economics is "non-operational" since its prescriptions are too often based on limited or false assumptions, it by-passes critical operational problems, and it ascribes too great validity to its statistical tests.

> By Barbara Bergmann (Eastern Economic Association, 1974), who prefaced her plea for more microsimulation to incorporate "realistically messy information" in our economic data base with a few roundhouse swings at the economics profession and the pointed observation that instead of studying the real nature of decision making, we typically rush to make assumptions "whose purpose in life is to let the theorem emerge, all neat and provable."

Another favorite line of criticism and attack focusses on the implicit value premises of conventional economics. Gunnar Myrdal and Robert Heilbroner chide us for concealing the value judgments that inevitably enter into our selection of problems for study, choice of approach, definition of concepts, and even gathering of data. So a "value-free" economics is an illusion—they urge economists to specify their values and thus avoid biases and make research more realistic.

Radical economists simply reject the whole value system of conventional economics—as they see it, the neoclassical paradigm in its very bone and marrow enthrones acquisitiveness and enshrines the existing order. Paul Sweezy accuses mainstream economists "of hiding the facts, of making the uncontrollable appear under control, of rationalizing a system which condemns hundreds of millions of people to lives of despair and starvation. . . ."

Inflation is the latest source of critical volleys, and I will get to these in due course. Meanwhile, the sampler of economic masochism I have already provided should serve as ample insurance against complacency or smugness in considering "what's right with economics." At the same time, it strongly suggests that economics, more than any other social science, is afflicted with the common scold.

I recognize that such a quick sampling and cryptic quotes, selected to highlight criticism, do a certain injustice to economics and to some of the quoted economists whose kindlier observations have been neglected in the process. But I am also aware that my litany omitted a number of familiar flaws, for example, our impounding of tastes and preferences in *ceteris paribus*; the shortcomings of the maximization principle in explaining consumer and producer behavior, especially in the short run; and our limited ability to bring the claims of future generations into our social utility functions.

Were I to serve as defense counsel for the profession on this wide variety of indictments, I would urge that we plead guilty or take the Fifth on some, take to the defense on others, and take offense at the rest. Having paid my respects to the critics, I intend no point-by-point evaluation or rebuttal. This has been ably undertaken by others.[1] Rather, my object is to gain a more balanced perspective by focussing on the quality, role, and contributions of economics, especially to public policy. In that undertaking, the first step is to examine the flank we expose to the public.

II. The Economist and the Public

When we turn from inside to outside critics, the focus changes. We may think, rightly, that freely confessing our weaknesses and airing our differences stimulate responses and adaptations that strengthen economics. Yet, wearing our purple hearts on our sleeves has its price. It nourishes the darkest suspicions about our art and supplies live ammunition to outside critics who have declared open season on economists. Witness the open sesame to the ed-op pages for such recent thrusts as Bergmann's assault on economists in general and Friedrich von Hayek's attack on Keynesians in particular. With everything from off-the-cuff phrases about being "caught with our parameters down" to tracts for the *Times*, we feed the hand that bites us.

This is not a plea to do our self-flagellating in secret or to mute our disputes and conflicts. Open controversies, openly arrived at, are part of the therapy that keeps our profession healthy. Rather, my plea is to the media and the opinion makers to understand that appearances are deceiving, that hard give-and-take is indeed a symbol of strength, and that our areas of agreement and consensus are vastly larger than our areas of difference.

On the first point, observers from other disciplines are often astonished at how hard economists go at each other, how readily they run the gauntlet of their colleagues' criticisms with no quarter asked and none given—and, with few exceptions, all this within the framework of professional respect and friendship. As Charles Frankel put it, unlike other social sciences, economics seems to have achieved "a working etiquette which allows people to disagree vigorously without engaging in recrimination about 'unscientific' or 'unprofessional' behavior" (quoted in Johnson (1973)).

What accounts for this? Part of it, one can unblushingly say, is simply that so many competent, tough, and rigorously trained minds have been drawn into economics in response not just to challenging policy problems but to the quantitative revolution since World War II. And part of it is that the participants can draw on a hard core of economic theory and methodology, together with a growing body of empirical knowledge, to provide standards for testing the validity (though not necessarily the relevance and reality) of ideas, analysis, and empirical findings. The result is not only a relentless intellectual policing of the profession that soon exposes the fool, the quack, and the charlatan, but a growing capacity "to participate in adversary debate over public policy issues without jeopardizing scientific integrity and freedom" (Johnson (1973)).

---

[1] Among those who have sprung to the defense with varying degrees of fervor are Harry Johnson (1968), Donald MacDougall, Charles Schultze (1972), Robert Solow (1970, 1971), and James Tobin (1973, 1974). For more general appraisals of the criticisms and the state of economics, see Blackman and Nancy Ruggles.

That brings me to the second point, the impression we give outsiders of a house divided, not to say splintered. It is worth reminding ourselves and our critics of several factors that drive a wedge between image and reality.

One, instead of laying aside our differences and living contentedly together, we economists tend to lay aside our agreements and live contentiously together. We focus our private and public debates on unsolved policy problems, tough analytical nuts, and issues on which we have rival theories, contradictory evidence, or strong ideological differences. Just as these are the questions that intrigue us, they are the ones that attract the attention of press and public. What we know—and they may not—is that beneath the visible tip of disagreement and rivalry lies no huge iceberg of divisiveness.

Two, it is only occasionally that our areas of consensus are brought to the surface in a newsworthy way. One such occasion was the White House "summit conference" on inflation last September. Two dozen leading economists from across a wide spectrum of American economics (not wide enough, the radicals would say) signed a statement which called on the President and Congress to eliminate twenty-two restrictive laws and practices that inhibit competition, inflate costs, and prop up prices. Only a tiny minority held out (if any minority that includes Galbraith can be called "tiny"). Even more striking, in a sense, was that while the customary and largely ideological clashes among, say, Galbraith, Milton Friedman, and Paul Samuelson caught the public eye, the real story lay in the minimal dissent among the participants on (a) the forecast of a soggy or sagging economy, (b) the urgency of providing relief to the victims of inflation and the casualties of recession, (c) the need to ease monetary restraint, (d) the small anti-inflationary payoff on moderate ($5 to $10 billion) budget cuts, and (e) the advisability of resisting popular demands for reimposing full-scale wage and price controls.

Three, even where disagreement flourishes—most visibly, perhaps, between Keynesians and monetarists—the public may not discern that the analytical and empirical ties that bind us are far stronger than the forces that divide us. Our controversies take place within the context of basic consensus on the nature and methods of economic theory and inquiry, on the content of the disagreement, and on the kinds of tests that may one day resolve the conflict. "Such disagreement within agreement lies at the heart of the process of normal development of a science" (Benjamin Ward, p. 12).

Four, much of what the public perceives as a clash of economic concepts and findings is in fact a clash of ideology and values. Given the way technical economics and ethical preferences are packaged in policy debates (and given our lapses in identifying which is which), this is hardly surprising. Thus, whoever opens the package labeled "monetarist" typically finds not just money supply in full flower, but a dedication to minimum government intervention, small budgets, reliance on rules rather than authority, and price stability. Contrasting correlations appear in the Keynesian package. So outsiders can be excused for slipping into the fallacy of association and attributing the split to our unresolved analytical conflicts rather than to divergent evaluations of social priorities and competing philosophies of government. These associational chains are not linked together by any inexorable logic—in part, they seem to be an accident of birth as in the case of the Chicago twins of monetarism and *laissez-faire* rules. A belief in the supremacy of monetary over fiscal tools could quite logically go hand-in-hand with avid interventionism. But this escapes the jaundiced eye of the outside observer, who takes the ideological lineup as further evidence that economics is riven to its core.

Five, there is an ironic but substantial inverse correlation between the degree of consensus among economists and the degree of public acceptance of their findings. Thus, in the macro-economic sphere of stabilization policy, where debate and disputes among economists flourish, their imprint on public policy is undeniable. But in the considerably more peaceful realm of microeconomics and allocative efficiency —where a reliable analytical apparatus coupled with solid quantitative work, especially on costs and benefits, has led the great majority of disinterested economists to an agreed diagnosis and prescription— the policy box score shows few hits, fewer runs, and lots of runners left on base. Economists widely, in some cases almost uniformly, favor tougher antitrust policy, freer trade, deregulation of transportation, pollution taxes in place of most prohibitions, and tax reform to remove income tax shelters. They oppose fair trade laws, restrictive labor and management practices, distortive zoning laws and building codes, import quotas, ceilings on interest rates, maritime subsidies, and pure (or impure) pork barrel projects.

Granted, the diffuse and inchoate consumer interest has been no match for the sharply focussed, articulate, and well-financed efforts of producer groups. But the economist is beginning to pick up some allies. Public interest groups are increasingly giving focus and force to the consumer and general public interest. And the march of events is providing some windfalls: Among the apples that have dropped in our laps are flexible exchange rates, the dethroning of agricultural price supports, inroads on import quotas, and moves to end percentage depletion. Under the pressure of virulent inflation, government actions that erode productivity and boost costs and prices are being subjected to new and searching scrutiny. So perhaps, on these micro-economic issues where economists sing in reasonably close harmony, the outside world will no longer quite tune us out. In macro-economic policy, where cacaphony prevails, we can be sure that the world will tune us in.

It may also be useful to draw attention —especially the attention of those who interpret us to the public—to certain other misperceptions and roadblocks that thwart good economics and tend to put economists in bad repute.

First, much of our economic analysis and the uncommon sense growing out of it fly in the face of "common sense," for example: that budget deficits need not spell inflation, nor national debt a burden on our grandchildren; that thriftiness can be a mixed virtue; that while exploding oil prices *in*flate costs, they *de*flate demand; that in an overheated economy, greater taxes can be the lesser evil; and so on. Behind every false dictate of common sense lies a primitive and misbegotten economic theory—and for most of our pains to correct it, we can expect to get the back of everyman's hand.

Second, a related cross to bear can be characterized by Kermit Gordon's apt phrase, "virtue is so much easier when duty and self-interest coincide." Not only does that foredoom action on many microeconomic fronts, as already noted, but it puts roadblocks in the path of efforts to make fiscal policy a two-way street. For forty years, Congress has enacted major tax increases only under the whiplash of war. The resulting reliance on tight money to fight peacetime excess demand, coupled with expansionary fiscal policy to fight recession and slack, have had an unmistakeable ratchet effect that has tilted the system toward tighter money and easier budgets. (Small wonder, by the way, that many economists and policy makers are unwilling to give up, via indexing, the increases in effective income tax rates "legislated" by inflation.)

Third, the public sees economists as the bearers of hard and unpalatable truths. And often we are, by the very nature of our sometimes dismal discipline. Except when idle resources can be put to work or productivity increased, our message is the stern one of tradeoffs, benefits at a cost, and no one-dimensional daydreaming. Even worse, at times economics has to bring the bad tidings that for some problems there are no satisfactory solutions. For some thirty years, we have warned that full employment, price stability, and full freedom of economic choice cannot coexist in a world of strongly organized producer groups. More recently, economic analysis has brought home the unromantic truth that failure to cure some of our social ills traces less to a failure of will, or "right-wing villains," or a calloused "establishment," or powerlessness of the people, than it does to the prosaic facts that the problems are tough and complex and the goals we seek may be irreconcilable—in short, trace more to conflicts in our national objectives than to conflicts among social groups. Welfare reform is a case in point: no solution can simultaneously provide a decent minimum income for all, preserve work incentives, cut no one's benefits, and avoid huge budget costs (see Schultze (1972); Rivlin (1973)). *We* may view such work as a contribution to straight thinking and rational choice. Our critics are more likely to view it, at worst, as a counsel of defeat (which it is not) or at best a counsel of inescapable compromise (which it is).

Since the foregoing misperceptions and roadblocks thwart the translation of good economics into good policy, one could justify, in cold cost-benefit terms, a sizeable investment to overcome or reduce them. The most obvious implication is that the country needs to invest more in formal economic education at all levels. But an equally pressing need is for economists to invest more of their time and effort in making themselves understood to the public and policy maker—and that in turn requires recognition of this skill in the academic reward system. This might serve as a useful antidote to the influence of mathematics and econometrics which, while heightening the precision of professional thinking and internal communication, have apparently dulled the appetite and eroded the facility to communicate with the public in intelligible English prose.

In a very real sense, this confronts the press with an unusual opportunity and challenge, perhaps even a responsibility, to serve as a translator and interpreter of economics and its offerings. But believing (probably rightly) that their readers and listeners prefer to hear of fights and failures, crises and controversies, rather than of quiet contributions and consensus, the conventional or electronic press is not very likely to rise to this challenge. So it is still up to economists. But I must not carry this too far. Just as I am eschewing any Cassandra-like pronouncements tonight, so I have promised myself to suppress the oracular and even the avuncular (Dutch-type) mood evoked by these occasions. So I shall press on.

III. Standards of Judgment

From the foregoing, it is evident that I feel, first, that economists have gone beyond beguiling humility and welcome self-criticism to the point of almost neurotic self-rebuke, and second, that press and public have all too lustily taken up the cry—in part taking us at our word, in part misinterpreting us, and in part reflecting their belief that, after the high promise of the 1960's, we have failed them in not foreseeing and forestalling the crises of the 1970's: stagflation, energy shortage, and the environment.

In my quest for a more balanced perspective on the state of economics, the next task is to set up some standards for judging the quality and performance of economists. Since we have developed no measures of output or allocative efficiency, no capital-output or cost-benefit ratios, for the economics "industry," I will have to fall back on more subjective and less quantitative measures in judging its quality and contributions.

My mixed bag of criteria includes (1) the quality of inputs; (2) the demands for our services; (3) as a proxy for a measure of outputs, the record of accomplishment in a given field (public finance); (4) finally, the cruelest test, our handling of the economics of inflation.

The *potential* of economics for informing and improving public policy depends on the stock of human capital, technology and tools at its command. Here, economics has no difficulty in holding its head high, especially in terms of the striking advances of the past three or four decades.

Harry Johnson may be a trifle extravagant in his assessment that the United States now has "perhaps fifty economic departments of an average quality comparable to the average quality of the four or five best departments in the whole world in the pre-World War II period . . ." (1973), but only a trifle. Another attest to professional quality, already referred to, might be put this way: Show me another field that has enough inner strength to confess so much remaining weakness (and to carry on so much open controversy). Humility where we have things to be humble about (and we do) is a becoming trait. But coupling it with pride where we have things to be genuinely proud about is hardly a deadly sin.

Accompanying the growth in the quantity and quality of economic brainpower have been striking advances in the techniques and tools with which economists work. For this audience, I can speak in shorthand about the strengthened analytical base of micro- and macroeconomics; the methodological revolution that moved us from the rationalist-historical approach into the age of quantification, with its insistence on systematic measurement of the shapes of economic functions and empirical testing of hypotheses and its use of econometrics and simulation techniques (with a powerful assist from the computer); and such conceptual advances as those in the economics of human capital, of cost-benefit relations, of uncertainty, of control, of transactions and information costs, of "second best," and of the allocation of time.

In normative economics and the analysis of value-laden social problems, new frontiers in the study of economic behavior are being opened up by survey research techniques (especially by the Michigan Survey Research Center), by efforts to measure nonmarket benefits or values (especially by the National Bureau of Economic Research) and by "controlled" social experimentation (for example, by the Brookings Institution and the University of Wisconsin Institute for Research on Poverty). These newer tools and the institutions that nurture them constitute part of the rich and expanding resources of economics.

Economics can also draw on a broad data base, especially in federal statistics. But here, the quantity, timeliness, and even the quality of the data are not keeping pace with either the problems requiring analysis or the capacity of our quantitative techniques. Responding to policy needs and mounting self-criticism, the profession has opened many new fronts in the search for realistic micro-data to link up with macro-data, for cross-section data to help overcome the curse of collinearity in time-series analysis, and for custom-built data developed by survey and experimental techniques.

That the human, analytical, and quantitative resources of economics provide a huge potential for solving problems seems undeniable. That more of these powerful resources than ever before are being put at the disposal of economic policy makers also seems undeniable. What we do not know is what proportion is being misdirected into arid puzzles, sterile proofs, and recreational mathematics while the world's pressing economic and social problems go begging for answers. Here, we can only match one observer's impression against another's. The profession itself has not come to grips with *this* question of allocative efficiency.

A second test in appraising the state of economics, one not unknown to economics, is that of the market place. This takes several forms, none very robust, but none trivial. The first is the upsurge in enrollments in economics courses, especially in introductory economics, that has occurred in the academic years 1973–74 and 1974–75. The second is the oft-reported high ranking of economists' salaries in business, government, and academic life. A third is the strong and growing demand for economists' inputs into the policy-making process—either as staff members or as expert witnesses for congressional committees, individual congressmen, and the executive branch.

With students, business, and government beating a path to our door, we can infer that something must be right with economics, or wrong with the economy, or both. Either we are building a better mousetrap or there are more and bigger mice threatening our customers. Perhaps it is simply that we have the only mousetraps in town.

But there must be more to it than that. Take the policy maker, for example. What he finds congenial is that he can hand an economist a problem—relating to changes in taxation, regulations, budget proposals, pollution control, poverty, social security, public service jobs, gasoline taxes, oil prices, and so on—and be reasonably sure of getting a useful appraisal of alternative paths to his objectives, of costs and benefits, and of distributional, allocative, and stabilization impacts. Many of these judgments will come with orders of magnitude or reasonably precise numbers attached. He may not trust our *GNP* forecasts, but he has come to respect our hardheaded analysis and numbers on the myriad problems of economic choice with which he is faced.

It seems fair to draw another inference: notwithstanding the current wave of self-criticism and public criticism, even lampooning, of economists and despite our highly visible public debates and highly vulnerable participation in policy making, economics continues to maintain its standing as a science. Signs of a reported crisis of public confidence or of a "recession of self-confidence" are few and far between.

Reports of the demise of our discipline are grossly exaggerated.

IV. The "Outputs" of Public Economics

Having considered some indicators of the quality of our inputs and of the revealed preferences for our outputs, let me continue this exercise in casual (and congenial) empiricism by taking an unscientific but not unrepresentative sample of the outputs of economics, especially those bearing on policy. For this purpose, I draw on my chosen field of public finance, or public economics, to illustrate the telling conceptual and empirical advances of economics in recent decades and the resulting enrichment of its offerings to the policy maker. Indeed, such an appraisal offers so many healthy antidotes to "what's wrong with economics" that I was tempted to devote my whole discourse to it tonight. But I resisted the temptation because, first, much of it has already been done in carefully documented depth in survey volumes by Brookings and the National Bureau;[2] second, I figured it might test your patience and mine; and third, it would have left no room for a confrontation with inflation. So I offer instead a miniaturized assessment of the achievements of public economics as viewed through the policy prism.

Public Expenditures

Consider first the striking contributions economics has made in the past generation to clear thinking and better informed decisions on public expenditures. Partly, this reflects advances in economic science, for example, in the theory of public goods and human capital, and partly, creative new applications of the economist's characteristic way of looking at problems of choice, namely, through the lens of opportunity costs, benefits, and alternative paths to a stated goal.

Economics can offer much more concrete guidance on efficient ways of allocating resources to achieve stated governmental objectives than it can on what the public-private sector division of resources should be. That may be a good thing in that presidents and congressmen view the fixing of goals for public health, housing, welfare, and the like as what *they* were elected for, yet at the same time seek, or at least accept, economic guidance on the choice among competing methods of achieving these goals.

Nonetheless, rapid progress in the theory of public goods since the appearance of the Samuelson classic on "The Pure Theory of Public Expenditures" just twenty years ago has vastly improved on the simplistic theory it replaced. It has facilitated straight thinking on the derivation of conditions for efficient public-sector allocations from private evaluations and on the articulation of social priorities through the political process.

Interwoven with the newer thinking about public goods has been a resurgent interest in externalities or spillover effects. In a sense, the pure collective good is a case of total externality—all of its benefits are external and nonmarketable since nobody can be excluded from them. That may clarify thinking but gives little policy guidance.

Yet, the externality concept translates into hard-headed policy advice in such disparate areas as pollution, federal aid, and the law. When pollution became a national concern, economists quickly drew on their tool kit to develop proposals for antipollution taxes (within the context of target air and water quality standards). Tax penalties of so much per unit would put price tags on use of the public's air and water, thus internalizing external costs and using market incentives to accomplish depollution rather than relying on the less efficient route of regulation.

---

[2] See Alan Blinder and Robert Solow et al. This is the capstone volume of the Brookings Studies on Government Finance, directed by Joseph A. Pechman, which has produced 35 books in the past decade. See also Carl Shoup et al. This was one of several survey volumes under the general heading, *Economic Research: Retrospect and Prospect*, based on the Bureau's Fiftieth Anniversary Colloquia.

When local governments supply education and public health services to a mobile population, many of the benefits spill over to other units. An important rationale for federal grants flows from these externalities, namely, that to get local units to produce enough education and health service to achieve a national, not just a local, cost-benefit optimum requires conditional grants from the federal purse.

Further, since externalities in the form of damage to third parties lies at the heart of many problems in legal justice, economics is able to make an important contribution in this area.

When we turn to the empirical outputs that are now illuminating problems of public choice, we find the past decade bristling with new thinking, new techniques, and new measurements. These offer the decision maker important new guides in the selection and evaluation of government programs and new insights into alternative systems of delivering government services:

> Measurement of cost-benefit ratios has developed from the early metrics of water projects into, first, a sophisticated cost-benefit calculus for tangible investments like dams, roads, pollution-control projects and, second, cost-benefit estimates for intangible investments in human brainpower, skills, and health. Shadow pricing has been one of the useful tools in this connection. Cost-benefit analysis, even with its limits of quantification and its inability to shed light on distributional and value questions, is an important aid to informed decisions.
>
> A related advance is the development of new and tougher standards for judging government programs. The former criteria centered on the question: Is the program put into effect quickly and with high fidelity to the congressional intent? Now, the accountability question is: Does it deliver the goods? Does it accomplish the objectives? Inputs used to be stressed—if they conformed with the intent of the legislation, they tended to be judged a success. But now we try to measure outputs, a tougher and more elusive standard. (The parallel with judging the performance of economics and economists is painfully obvious.) Antipoverty programs, which were among the first to be evaluated by these stringent standards, seem to have borne the brunt of the evaluation boom. By the old inputs standard, a program like Head Start would have fared much better.
>
> The reach of cost-benefit analysis will be lengthened if a broad range of new research efforts in nonmarket sectors of economic activity pays off. I refer not only to the exciting work on measurement of the returns on investments in human capital (T. W. Schultz), but to efforts to measure the output of the medical industry, to measure the relations between crime and punishment, and to measure the value of nonmarket economic activity conducted *within* firms and households.
>
> The new technique of controlled social experimentation on proposed welfare and housing measures, health insurance, and education vouchers is yielding important insights. As a result of experiments on negative income taxation, for example, the equity versus efficiency, or equality versus incentives, controversy will never be conducted in a vacuum

again (Rivlin (1973)). In spite of some limitations, the New Jersey experiment yielded strong evidence that fears of fatal incentive effects of a negative income tax were grossly overblown.

Another focus of fruitful thinking relates to alternative strategies for delivering social services. The in-cash versus in-kind choice is a basic one. Economists are predisposed toward the in-cash approach on grounds that one can generally depend on people to follow their own best interests. But there are significant exceptions where consumer sovereignty is limited or specific goods externalities exist or some explicit social values take priority.

Out of economics also comes the attempt to develop "market analogs" to serve as substitutes for market incentives in reconciling public with private interests, decentralized individual decisions with social goals. Pollution taxes are a case in point. Performance standards for teacher pay would be another. Putting medical insurance programs on an efficiency-based reimbursement basis would be a third. The big gap is in the redesign of incentives and institutions to guide decentralized government decision making more systematically toward the aims of our social programs (Schultze (1971)). Thus far, the government, like the economics profession, is largely in the dark about its own production function.

Taxation

What strikes an old public finance functionary as forcibly as any change in the field of public finance is the way in which modern thinking has knocked the props out from under the neat and primitive theories of tax incidence of a generation ago. The property tax provides a particularly instructive case in point. The textbooks of the 1930's and 1940's told us confidently that the tax on land (fixed supply) was capitalized and on dwellings (supply-responsive) fell like an excise tax on the occupant, the consumer of housing services. The policy lesson was clear: Given the declining proportion of income spent on housing services as income rises, the tax was hopelessly regressive. Today? It is recognized that the old incidence analysis was wrong, even on its own terms.

The modern theory of incidence (defined as the impact on distribution of private real income) draws on general equilibrium theory, distinguishes between sources-of-income and uses-of-income effects, and disentangles the concepts of specific, differential, and balanced-budget incidence. The resulting analysis indicates that much of the aggregate burden of the property tax falls on owners of capital and hence tends to be progressive—and this progressivity is enhanced by the particular "excise-type" effects of this tax (Henry Aaron). In short, error has been exposed and though the debate is not over, we are now in transit toward truth.[3] It is hard to put down the knee-jerk reaction that prefixes "property tax" with "regressive." And it will take some time before policy makers accept the proposition that, at the very least, the property tax is now in the unexpected position of "innocent until proved guilty." But the implications for policy are profound.

Economists have long been useful and influential contributors to the design of the federal tax structure and of particular taxes. Again, elementary concepts we now take for granted—for example, horizontal versus vertical equity, Richard Musgrave's three branches of distribution, allocation, and stabilization, the lagged effect of tax changes, and automatic versus discretionary tax changes were not even part of our vocabulary in the pre-World War II period. Yet, all of these are now factored into our economic advice on taxation.

---

[3] Those who view decisions to locate in a particular community as a conscious choice of one particular bundle of public services over others conclude that the property tax on housing is a benefit tax, a payment for benefits received.

Even more directly impinging on policy are the empirical advances. One thinks of searching studies of particular taxes and tax components (especially in the Brookings Studies on Government Finance), and of the relentless identifying and quantifying of federal income tax preferences or "loopholes." Much of the thrust of economists' recent work on these "tax expenditures" has been (a) to identify the beneficiaries and specify the size of the government subsidies provided in the form of preferential tax treatment, (b) to define the inequities, both horizontal and vertical, that they create, and (c) to estimate the distortions in resource flows caused by preferential treatment of oil and gas, housing, real estate partnerships, and the like, and measure the resulting welfare loss. Though the congressional response has been slow and halting, progress has been made along the lines plotted by economists, and a solid base has been laid for the further tax reform that is surely coming.

Out of the countless other advances, one stands out, namely, the highly informative work done on the distributional impacts of taxation with the aid of the powerful tool of micro-unit data files (for example, the MERGE file developed by Joseph Pechman and Benjamin Okner). Such micro-unit files are a new-generation statistical missile, MIRVed so that they can simultaneously hit multiple revenue-estimating and burden-distribution targets. With their help, for example, economists have measured the growing burden of income, payroll, and consumption taxes on the lower income groups and developed techniques for removing them—most recently, in the context of the impact of inflation on the same groups.

One should add that if revolution rather than reform becomes the order of the day in the federal tax structure, the economist is ever ready with reasonably sophisticated analytics and a fair amount of empirical information on such major alternatives as a value-added tax, a progressive expenditure tax, and a net-worth tax. One of the next stages in tax research, a highly complex one, will be the general equilibrium analysis of such sweeping changes in the tax system as, say, the substitution of a value-added tax for the corporate income tax or for part of the payroll tax. Or, if stimulation of private saving becomes a compelling objective, perhaps the substitution of an expenditure tax for part of the income tax will become a live issue. The skills of the economist will be front and center in any such redesign of the tax system.

The negative income tax story is relevant here. The concept and its rudimentary principles were developed and discussed among economists in the early 1940's. Some of us were already using it as a teaching device in the mid-1940's. A quarter-century after its origin, it became the basis for the Family Assistance Plan developed by Mr. Nixon's economists. And a more limited version of the plan seems again to be rustling in the leaves.

Fiscal Policy

In the domain of fiscal policy, it is harder to answer the question, "What have you economists done for us lately?" with a sparkling array of examples. Much of the theoretical ferment in this field is associated with the flowering of Keynesian macroeconomics in the late 1930's and 1940's, the very period when the microeconomics of tax incidence and public expenditures languished.

Conceptual advances have continued throughout the past twenty-five years, but they have been more in the nature of a fleshing out and consolidation of the original breakthroughs with the aid of the powerful tools of mathematics and econometrics. Multiplier analysis, for example, has moved from the theoretical realm into large computer models of the economy—with the tax cut of 1964 and the surtax of 1968 providing empirical grist for the mill. While the models differ on the exact value of the multiplier, "a fiscal policy planner will not often be led astray if he uses a multiplier of 2" for government spending (see Blinder and Solow).

Coupled with multiplier studies is the even more subtle study of the structure of the "outside lags," of the timing of responses in the economy to changes in fiscal policy. Though the empirical efforts and debates go on apace, the behavior of the cumulative multipliers in a clutch of economic models suggests that for any given change in fiscal policy, "at least 75 percent, probably much more, of the ultimate effect is felt within the first year after the initiation of the policy" (Blinder and Solow). Although intractable questions remain concerning investment responses to fiscal policy changes, enough has been learned about aggregate demand responses to provide two broad generalizations about fiscal policy:

> One, the conditions for intelligent fiscal policy are met if economic forecasting can answer two not-very-exacting questions: Do projected economic conditions in the ensuing six to nine months call for restraint or stimulus? Is the required dosage large or small?
>
> Two, given the limited margin for error in a high-employment economy, it is better to rely on many smaller monetary-fiscal moves than a few large ones.

Implicit in these two generalizations is a third one: Given both the internal shifts and the external shocks with which stabilization policy has to cope, a discretionary policy that makes efficient use of feedback information will be more effective than an automatic policy that locks in on fixed fiscal and monetary targets.

Development of a simplified measure of fiscal impact revolving around the "full employment surplus" (*FES*) concept is another example of the typical process by which economists expose error, develop approximations of truth, but continue the vigorous debate on further improvements. First, policy makers had to be weaned away from the annually balanced budget and the cyclically balanced budget as policy targets and from actual deficits or surpluses (especially in budgets other than the national income accounts budget) as measures of budget stimulus or restriction. It was not easy. It took almost a quarter of a century before a Democratic president was converted (in 1961) and another decade to capture a Republican White House.

But success on the policy frontier has its own pitfalls, both political and economic. What was intended as a measure of policy was instead taken as a goal, namely, a balanced budget at full employment, a "self-fulfilling prophecy" as the Nixon Administration called it. This erroneously implied that the fiscal target should remain fixed regardless of changes in monetary policy and significant shifts in private demand, for example, a plant-and-equipment boom. Apart from trying to correct such misconceptions, economists have had to wrestle with the problem of the overstatement of the full employment surplus when inflation expands revenues faster than expenditures, not to mention the problem of weighting for differing multipliers if tax or expenditure components change sharply. In brief, the advances over the bad old days of the annually balanced budget are enormous, but economists are aware of the limitations of the *FES* measure and are struggling to resolve them.

Just as economics relegated erroneous budget concepts to the dustbin, so it has cast a shadow over such former favorites (of mine, among others) as federal capital budgeting and the "shelf of public works." The initial enthusiasm for the capital budget concept (in the context of a Congress seeking to balance the budget annually) was dispelled by second-thought analysis showing that (a) it rested on some faulty parallels with private finance, (b) the implicit fiscal policy rule of always financing capital projects by borrowing is in error, and (c) it would bias government capital spending toward bricks and mortar instead of brainpower and people. In the public works case, the concept ran afoul the findings of prosaic economic research: recent studies show that the public works program launched in 1963 to speed recovery was far from completed before excess demand overtook us in the 1966–69 period. This is not to rule out the use of certain types of "public works" that are nimble on their feet, such as road and forest maintenance work, for stabilization purposes. Nor does it rule out speeding up or delaying the launching of projects that are to be undertaken for sound cost-benefit reasons in any event. But it is fair warning not to expect very much stabilization help from the public works sector (not to be confused with public service employment).

In the conscious use of taxes for stabilization purposes, the huge 1964 income tax cut delivered economic expansion and a balanced budget on schedule without inflation by mid-1965, just before the Vietnam escalation struck the economy. The temporary 1968 surtax, buffeted by powerful demand forces and monetary easing, left a more ambiguous econometric trail. Subsequent fiscal policy thinking emphasizes the advantages of temporary tax changes that embody not just income effects but intertemporal substitution effects. For example, lowering the prices of investment goods in a recession via a clearly temporary increase in the investment credit, or temporary cuts in consumption taxes on durable goods (or lacking these, temporary purchase subsidies), would constitute a powerful incentive to purchase those goods before the price went up again.

Further work is needed to measure the cost-push effects of anti-inflationary tax increases that offset part of their demand-damping effect. In recession, the cost-easing and demand-push effects work in happy harmony. They work at cross purposes in tax increases (though not in expenditure cuts) to curb inflation. The question of how large the offsetting cost-push effects, or aggregate supply effects, may be, is unresolved. In a high-inflation economy, this is a serious gap in our fiscal policy knowledge.

Other Aspects

This kaleidoscope of contributions, long as it is, leaves out a whole string of developments in budget concepts, techniques, and processes—efforts that were crowned by the congressional budget reforms recently put into effect. Much of the guidance and momentum for these reforms was provided by economic analysis and by a succession of five economist-budget directors throughout the 1960's. Also omitted is the conceptual work on the economics of the bureaucratic process, of how government works. Other omissions include the rebirth of interest and great advances in the economics of state-local finance, the rapid growth of the important new field of urban economics—with its contributions to regional economics, location research, and analysis of the city as an economic system—and the enriched economics of fiscal federalism. I have even eschewed an assessment of revenue sharing, the rationale and form of which were developed by economists. With little imperialism, economists can also cite the firm quantitative evidence being developed to demonstrate the adverse economic effects of many public regulatory activities.[4]

For all the advances, the agenda of unresolved conceptual questions and unfinished empirical business is huge. But even this truncated review of progress and current output in public economics makes clear that the contributions of recent decades have enormously enriched this field not only conceptually but as a source of hard practical advice to decision makers who want to shape a better tax system, do justice to the poor, improve social programs, reform budget procedures, fight unemployment, and so on. And in the process, the frontiers of normative economics, both theoretical and empirical, have been pushed out into the areas of education, health, racism, crime, family behavior, and even political behavior.

As a result, we have plunged ever deeper into the realm of values. Not that it was a value-free inquiry to ask the traditional questions about the effect of a given policy on material output. But surely the testing of policies by the costs they incur and how effective they are in meeting some generally accepted criteria of social welfare or general welfare involves economics directly in value and distributional problems. And it enables economics to say important things on social policy issues within the framework of the conventional economic paradigm and with rigor of the non-mortis variety.

We are becoming interdisciplinary in spite of ourselves. When *we* do it, of course, we don't think of it as cross-sterilization of disciplines. But here is an area where modesty becomes us. For if we confine ourselves too narrowly to economics, we are far too likely to attribute to economic variables the behavior and results that are really a response to social variables. Fearing just that, one observer has been unkind enough to suggest that we ought to stick to inflation problems where we *all* know what to say.

V. The Economist and Inflation

Inflation may no longer be "Public Enemy Number One" now that severe recession is upon us, but it is surely "Economists' Enemy Number One." Among the charges of, by, and against economists that have been touched off by double-digit inflation and reported in the public prints are these:

> Economists have confessed (I plead guilty) that 1973 was "the year of infamy in inflation forecasting" and, as already noted, that "we were caught with our parameters down."
>
> Aaron Gordon puts it more explicitly when he says that "the forecasters fell flat on their faces in predicting price changes because they didn't have any way of estimating sectoral supply scarcity" and adds that we have not "even started to develop a theory of aggregate supply."
>
> Leontief scolds macroeconomists more generally: "There is a lot of fancy methodology, but the macroeconomists get indigestion if you give them facts."

---

[4] As an example of the "Age of Quantification," George Stigler cites the sea of studies on regulatory practices and their costs and benefits in the past dozen years, where there was a vacuum before. He notes that thirty-six "quantitative studies of effects of laws" were reported in two journals alone during this period, the *Journal of Law and Economics* and the *Journal of Political Economy*. These are promoting a broader consensus within the profession, informing decision makers, and posing challenges that will make policy failures easier to identify. (Personal correspondence.)

> We are reminded *ad nauseam* that the "new economists" of the 1960's had promised to fine-tune inflation out of their full employment economy (a clear-cut triumph of caricature over fact since Keynesians time and again warned of precisely the opposite danger).

> Myrdal and Heilbroner have pointed to stagflation as Exhibit A that economists typically lag rather than lead their targets, that being "behind its time" is "the regular methodological weakness of establishment economics."

> Von Hayek recently reentered the fray to lay the blame for worldwide inflation squarely at the door of economists, particularly those "who have embraced the teachings of Lord Keynes."

Apart from the charge that Keynesian economists have *caused* inflation (which is much like saying that the cause of forest fires is trees), the bill of particulars against macroeconomics runs something like this: First, it did not forewarn the body politic that it would have to pay such a high price in endemic inflation for the attainment of high employment. Second, its progress in solving some important puzzles of endemic inflation relating, for example, to the Phillips curve, wage inflation, expectations, and uncertainty is much too slow. Third, there is no articulated general theory of inflation as such. Fourth, economists failed to foresee the 1973–74 epidemic inflation because their forecasting models lacked the central supply and price parameters. Fifth, macroeconomics is helpless in the face of epidemic or external-shock inflation—indeed, it has not satisfactorily explained the coexistence of inflation and recession, or stagflation. Without attempting a point-by-point assessment of these complaints, I will touch on all of them in the following sympathetic interpretation of how economists are coping with inflation's tough analytic and empirical challenges.

Addressing myself for a moment to our reproachful public, let me simply say to them: "We never promised you a rose garden without thorns." Over most of the past thirty years, macroeconomists have warned again and again, first, that aggressive fiscal and monetary policy to manage aggregate demand was bound to generate inflationary pressures once the economy entered the full employment zone, and second, that while full employment spells inflation, recessions run into price and wage rigidities that thwart deflation, an asymmetry bound to produce a ratchet effect on the price level. Keynes himself foresaw the basic problem in his little book, *How to Pay for the War*, in 1940. Abba Lerner and William Beveridge also wrote of the problem in the early 1940's. And it has been discussed in the stabilization theory and policy literature, in congressional hearings, and in other policy forums ever since.

This country finally embraced activist fiscal policies for full employment in the 1960's, most explicitly in the 1964 tax cut. Following the canons of Keynesian economics, focussing on the economy's full employment potential as their target, and steadfastly rejecting a spate of "structural" explanations of unemployment, economists were at first alone in prescribing tax cuts as a tonic for the stagnant economy. Enacted early in 1964, the tax cut delivered the promised expansion and budget balance without inflation. By August 1965, when Vietnam escalation began, unemployment had been brought to 4.4 percent with only the faintest stirring of the inflationary beast (i.e., with consumer prices rising at less than a 2 percent annual rate).

In a very real sense, economists have been victims of their own success. Macroeconomic policy, capped by the tax cut, was the major force holding the postwar economy on a vastly higher plane than the prewar economy.[5] On one hand, the high employment, limited-recession economy forged with our macro-economic policy tools is indeed an inflation-prone economy—the formula for successful management of high-pressure prosperity is far more elusive than the formula for getting there. Yet on the other hand, success bred great expectations on the part of the public that economics could deliver prosperity without inflation and with ever-growing material gains in the bargain. The message got through that we had "harnessed the existing economics . . . to the purposes of prosperity, stability, and growth," and that as to the role of the tax cut in breaking old molds of thinking, "nothing succeeds like success" (Heller). *The Economist* unkindly corrected me: "Nothing *ex*ceeds like success."[6]

To be sure, critics and converts alike ignored our caveats that the goal of "prosperity without a price-wage spiral" had "eluded not only this country but all of its industrial partners in the free world," that "the margin for error diminishes as the economy reaches the treasured but treacherous area of full employment . . . ," and that "the 'new economics' promises no money-back guarantees against occasional slowdowns or even recessions" (Heller).

All too soon, Vietnam blew the economy off-course. Economists found that in the political arena fiscal policy was not a two-way street and that the much delayed surtax adopted in mid-1968 was no match for surging inflation. Nor was the induced recession of 1969–70. It took a combination of the 1971 shock therapy of tight wage-price controls and the stimulus of tax cuts to subdue inflation and energize expansion. It is worth noting that economists analyzed and projected the effects of this "new economic policy" with exceptional precision. That the tax cuts, coupled with controls and devaluation, would generate a surging expansion at very moderate rates of inflation in 1972 was widely and accurately forecast.

But the period from August 1971 to January 1973 was in the nature of a remission from the inflationary disease, clearly not a cure. The 1969–70 recession brought home the worsening problem of persistent inflation in the face of slowdown and recession. It presented new empirical puzzles for the analysts of the Phillips curve, wage equations, and expectational inflation. And it began to prompt the public mutterings that are being intensified by the 1974–75 stagflation: "All right, so you did not promise us a rose garden without thorns—but the thorns without the rose garden?"

Keenly aware of these problems, economists have long been at the drawing boards on this problem of endemic inflation. In a close parallel with research on cancer, economists are working on various pieces of the inflation puzzle and producing useful insights and guidance for policy purposes. But as economists, we would be the first to underscore that these puzzles are far from being fitted into an articulated and holistic theory of inflation. Inflationary analysis appears as an appendage to Keynesian and monetarist theories. But as yet, the Keynesian apparatus cannot tell us how any given change in aggregate demand is divided between changes in real output and changes in prices. Nor has monetarist theory unlocked the puzzle of how the effects of monetary changes are divided between output and price level changes. And no big breakthrough is in sight.

---

[5] As gauges of the contrast between prewar and postwar performance: unemployment averaged 18.8 percent in the decade of depression (1931–40) in contrast with 4.8 percent in the twenty-eight years since World War II; the prewar peak annual rate was 24.9 percent, the postwar peak was 6.8 percent. Annual real *GNP* dropped 30 percent from 1929 to 1933; since the war, mild declines have occurred only in three years (1949, 1954, and 1970), though 1974–75 may add two more. Consumer prices in 1940 were 18 percent below 1929; from 1948 to 1974, they increased 106 percent.

[6] Macroeconomists were not alone in their exuberance in the mid-1960's. On this rostrum a decade ago, George Stigler, after reviewing the great promises and early accomplishments of the "Quantitative Revolution in Economics," was moved to say, "I am convinced that economics is finally at the threshold of its Golden Age—nay, we already have one foot through the door. . . . Our expanding theoretical and empirical studies will inevitably and irresistibly enter into the subject of public policy, and we shall develop a body of knowledge essential to intelligent policy formulation. And then, quite frankly, I hope that we become the ornaments of democratic society whose opinions on economic policy shall prevail."

Does this mean that the economist has to stand mute in the meanwhile? Not at all. He is pushing ahead on the various pieces of basic research on the cancer of inflation and isolating and prescribing effectively for particular forms of the cancer even without having a complete explanation of the disease. Let me come back to the sustained and systematic research efforts on endemic inflation after examining the 1973–74 epidemic and the economist's responses to it. Since the epidemic is an over-layer on the endemic base, the distinctions won't be clear-cut—but they are nonetheless useful for viewing what the economist is able to contribute to policy.

The food-fuel price bulge generated over half of the 1973–74 inflation—and of economists' woes as well. Yet, it is asking a lot of economists to expect them to have foreseen that the oil cartel would quadruple oil prices, that the world would suffer widespread and successive crop failures, that the Peruvian anchovies would go into hiding, and that the Soviets would "solve our surplus grain problem" overnight.

Several unpleasant policy surprises also beset the inflation forecasters. First, just when a new rash of inflation was breaking out early in 1973, the reasonably effective Phase II controls were abruptly dropped in favor of the weak and ineffective Phase III. Second, six months later, after inflation had changed into a commodity-driven structural phenomenon involving a drastic readjustment of relative prices, the White House (to the pained surprise of economists inside and outside the administration) prescribed just the wrong medicine, a new wage-price freeze. A third policy surprise was that the dollar was allowed to sink like a stone: At its low point in the summer of 1973 (just before a substantial rebound), relative prices of imports had risen 10 percent in six months. About a quarter of the 1973 inflation has been attributed to these policy developments (see William Nordhaus and John Shoven).

It is worth noting that unexpected twists and turns of federal policy—which might be termed "internal shocks" in contrast with the "external shocks" of the food-fuel price explosion—are a continuing bane of the forecaster's existence. The about-face of the Federal Reserve in 1974 is another painful case in point. The sharp turn from ease to tightness in the first quarter of the year was a major factor in transforming prospects of recovery into recession in the second half of 1974. It is not quite clear why economists should be better at anticipating these shocks, especially the external ones, than society as a whole, or other professional specialists, or practical men of the world. Nothing in statistical methodology or economic science enables us to predict random shocks. What *can* be expected of us is that when they occur, we will spot them quickly, identify them, and analyze their significance for policy.

It is also worth remembering that democratic governments, by their nature, are pressure-responders rather than problem-anticipators. This carries two implications for political economists. On one hand, if an idea's time has not yet come, or if a problem has not yet become a crisis, the economist's call for action is likely to go unheeded. On the other, spotting emergent problems early can perhaps hasten an idea's time and alert the policy makers to impending danger.

Economists can more readily be faulted for being caught by surprise by the shortages of materials and primary processing capacity that caused the economy to bump against its ceiling sooner than expected and by the worldwide economic boom that put severe pressure on raw commodity supplies and prices. On the first point, we suffered both from information failure—the official capacity indexes simply did not reveal how close the economy was to its output ceilings—and from analytic limits. While identifying the causes, economists have been unable to pinpoint the relative significance of the shortfall of investment that began in the late 1960's, of underinvestment caused by price controls, of delays induced by environmental policies, and of the surge in foreign demand touched off by devaluation. However, I should add that the shortages problem is meat and drink for economists, and they are responding (especially in the energy field) with new analyses of price elasticities, investment needs, and the like. All of a sudden, price theory is back in vogue, and elasticities have replaced multipliers as the badge of a policy maker's *savoir faire.*

Delays in perceiving that the *U.S.* economic expansion was part of a worldwide upsurge can again be laid more to lack of an adequate information system than to any inability to understand the underlying principles. Still, a better sense of history and of the emerging worldwide imbalance between growing aspirations and growing incomes on one hand and inelastic resource supply and lagging technology on the other would have made us more conscious and cautious. We are considerably less likely to be caught by surprise in the future in view of the new worldwide data networks that are being developed by Project LINK at the University of Pennsylvania and by Otto Eckstein and his colleagues at Data Resources Incorporated (*DRI*).

Without absolving economists, one should apply this operational test: With proper foresight, would tighter monetary and budget policy have been able to damp inflation? It is worth recalling, first, that the full employment budget was making a swing of over $10 billion towards restraint between fiscal 1973 and fiscal 1974 (from a $2 billion deficit to a $10 billion surplus under the old 4 percent unemployment standard) and that monetary policy pushed interest rates into the double-digit region; second, that there was little that an aggregate demand squeeze could have done to push world commodity prices down. So the answer is clear: Even tougher fiscal and monetary policy would have had limited scope in holding inflation down.

This is not to deny that generating a larger full employment surplus would have been the prudent course in calendar 1973. But it is worth noting that to offset the food and fuel price explosions—which were triggered by forces largely immune to *U.S.* fiscal and monetary policy—would have required a ***reduction*** of 3 percent in all other prices. Such a target implies depression-inducing doses of fiscal and monetary restriction, an unthinkable "solution."

Looking toward the future, many economists draw the lesson not that one should keep the economy's motor idling, but rather that one should provide it with safety devices and heavy-duty shock absorbers, for example, stock-piling of foodstuffs, oil, and basic raw materials, careful tracking of commodity exports, distant early warning systems to spot shortages-in-the-making, and conservation and development measures to limit dependence on foreign raw materials cartels. In other words, it is a call for better planning, better data, and faster conversion of knowledge into policy.

Another criterion of economists' responses to inflationary shocks is how quickly they adapted (read, "disaggregated") their macromodels, large and small, to incorporate new supply and price parameters that had previously been judged of second or third order importance and hence relegated to Marshall's *ceteris paribus* pound. Some of the mongrel pups impounded there turned out to be full blooded huskies, for example, food prices, the exchange value of the dollar, oil and other raw material supplies and prices. At first most economists were slow and the big models sluggish in their responses. After all, for two decades prices had moved in tandem with wages, with a year-by-year percentage-point differential of $2\frac{3}{4} \pm 1$. So most models relied on wage trends, with some adjustment for productivity and capacity behavior, to give them a fix on price trends. Their eyes were on labor market indicators rather than commodity supplies, exchange rates, and the like. After some initial delays, the model builders scrambled to disaggregate, to build microelements into their macromodels. For example, *DRI* now has good stage-of-processing models that absorb the impacts of food and energy explosions. Price elasticities are being built into the macromodels to reflect the impact of massive relative price changes on the macrodimensions of the economy.

The whole experience reminds us of the role and limits of econometric forecasting models. First, the combination of computers, mathematics, and econometrics cannot produce the miracles that the uninitiated may expect of them—there is no way of replicating reality with its 3 million equations, all of them non-linear. Second, their indispensable function is to bring us closer to reality and help the mind manage the previously unmanageable—they permit us to release vastly more animals from the *ceteris paribus* pound than we could manage without these tools. Third, they have to be constantly adjusted to plug in common sense, adjust the length of the lags, and bring in new dimensions of the problem. Else, they will lock out things that a more judgmental approach would include, and will fail to respond quickly to changes in order of importance.

So the inflation-shock experience has brought home the need not just to watch supply but to watch *all* the pieces lest the model prevail over the mind, rather than having the model help the mind prevail over matter. The macro-stalactites have to reach toward the micro-stalagmites, and vice versa. I hope that metaphor is not a portent of the pace at which the advance toward macro-micro fusion will proceed.

Economists who use judgmental models have shown us how to be the master rather than the slave of the computer. A case in point was the early analysis (especially by George Perry) of the macro-impact of the oil price increases. A year ago, his work had already brought out the oil paradox—the *in*flation of costs and hence prices, leading to a *de*flation of aggregate demand—and had provided some estimates of both. The insight that some $15 to $20 billion of consumer purchasing power would be siphoned off into the hands of oil producers and royalty collectors without any early return to the economy in the form of demand for imports or investment goods had important implications for demand-management policy—implications that were ignored until severe recession was full upon us.[7]

---

[7] Late in 1974, Perry undertook a more searching econometric probe with the benefit of actual rather than projected oil price data and with the aid of the large

These important insights into the macro-economic policy implications of oil prices fit into the broader efforts of economists to disentangle the sources of the current inflation and identify the appropriate remedies. They differentiate among (1) excess demand, which had spent most of its force by early in 1974, (2) the price-wage-price spiral, which began to turn more rapidly in 1974, and (3) external-shock or special-sector inflation, in particular, the commodity-price surges that permeate the present inflation and account for its special character and ferocity.

The first responds rather readily to monetary-fiscal pressure, the second responds more reluctantly, and the third is highly resistant to the demand-management measures of any given country. For the second and especially the third types, therefore, high costs in unemployment and foregone output have to be incurred for small gains in curbing inflation. So the distinction is an instructive one for policy—even when the instructions are ignored. As we meet here tonight, the economic lessons that were so long ignored are being painfully driven home by severe recession and unemployment coupled with continuing inflation. A much-belated consensus that fiscal and monetary stimulus can now be undertaken with minimal inflationary risk is rapidly forming.

The economists' three-ply classification of inflation sources is also useful in driving home another point: In most *U.S.* inflations, consisting of the first two types, one person's price is another person's income, so that in spite of some reshuffling, there is no net loss in real income. Not so in 1973–75. Commodity inflation has transferred tens of billions of dollars of *real* income out of the pockets of urban consumers and wage earners into the hands of farmers and foreigners where it is beyond the reach of the collective bargaining process. From this, several important inferences can be drawn:

> Point for point, this inflation *cum* relative price changes is harsher in its impact than previous postwar inflations.
>
> In this "no-win" inflation, the wage earner's loss has not generally been the employer's gain; hence, if the wage "catch-up" process succeeds in recouping the *full* rise in the cost of living, much of the wage increase will pass through to prices and thereby give the wage-price spiral another self-defeating turn.
>
> It follows, as various economists urged throughout 1974, that tax cuts to bolster the real income of labor, if put in the context of a social contract, might well relieve some of the pressure for higher wages.

In this respect, today's situation contrasts rather sharply with the 1950–51 inflation when a similarly rapid run-up in world commodity prices was accompanied by a rapid rise in profit margins side-by-side with vigorous federal policies to boost capacity. The ensuing combination of ebbing world market prices and wage increases that could be granted without generating higher product prices resulted in a remarkable four-year period of price stability from 1952 to 1956.

A closely allied economic insight goes to the nature of the inflationary process. It

---

scale formal models. His analysis shows that the purchasing power loss had reached $37 billion (annual rate) by the third quarter of 1974 and that the rise in the deflator attributable to the oil price jump was 3.8 percent. His analysis embraced not only the real-income effect (the transfer of real income from consumers to producers), but also the monetary-policy effect (the reduction of the real value of the money stock and the rise in interest rates stemming from the highly inelastic short-run demand for petroleum products), the automobile-demand effect (higher saving), and the induced-inflation effect (the price-wage-price effect) of the oil price rise on the macroeconomy.

P954

explains in good part why inflation is so stubborn even in the face of overly restrictive monetary-fiscal policy and rapidly mounting unemployment and slack in the economy. It is the sharp run-up in *relative* prices of food, fuel, and imported goods—coupled with the downward rigidities of wages and prices—that is the key to most of our stagflationary malaise today.

These downward rigidities are a striking example of the way in which economic solutions create their own problems and move the economist relentlessly from one new frontier to another. Once macroeconomics gave governments the know-how and tools of modern demand-management to avoid depression, and once the public caught on that even recessions are essentially man-made—chiefly by That Man in the White House, whoever he is, together with the Congress and the Federal Reserve Board—it became part of the politics of survival to hold employment high and keep recessions in check. Absent the fears of mass unemployment and prolonged recession, the risks of not cutting prices and not accepting lower wages are minimized. Having put the Great Depression of the 1930's far behind us, will we therefore have to live with the Great Inflation of the 1970's?

Essentially, the economist answers that, given the ratchet behavior of wages and prices, the price level can only float upward to accommodate the massive relative price increases of oil, grains, certain raw materials, and imported goods. These sharp changes in the composition of supply touch off reverberating price increases throughout the economy as prices in the scarce-supply sectors become costs in the less-scarce ones. The reverberations go on —in substantial part independent of the state of aggregate demand and hence of monetary and fiscal policy—until the prices of the initiating goods have risen sufficiently farther than prices in general to accomplish the necessary realignment of relative prices. This is the process going on now. It takes time, but not forever. It has much to do with double-digit inflation, but it does not condemn us to Weimar Republic inflation.

Solow (1975) reminds us that the supply-shift phenomenon bears a close relationship to the demand-shift analysis of the creeping inflation of the mid-1950's. At that time, the parallel process was touched off by an investment boom that put excess demand pressures on capital goods industries even when there was no excess aggregate demand in the economy. Given the downward rigidity and cost-oriented nature of wages and prices in areas of excess market power, the price level had to float upward to accommodate those relative price changes (see Schultze (1959)).

John Dunlop and other economists have emphasized that there is a closely related phenomenon on the wage side known as "scale wages" or "wage relativities" or even a "just wage" (see Robert Hall and Michael Piore). If the relative wage scale is thrown out of kilter by an outsized wage settlement in one industry, the others will writhe, twist, and turn until the old relationships are reestablished. There is only one way the wage structure can move to accommodate this process: Up. Again, the process burns itself out only when a new equilibrium has been established on a higher plateau.

The policy implications of the supply-shift, demand-shift, and wage-shift insights are reasonably clear. One is the limited scope of repressive monetary-fiscal policy in coping with this process. Another is that the key to a successful wage-price policy for these circumstances is to establish and effectuate norms for the pace-setters and thus thwart the wage-wage and price-price spirals and the interacting wage-price spiral. Once the process is launched, the role of a wage-price watchdog with teeth would be to see to it that the adjustment process is a limited and straightforward one, not a leapfrogging sequence that will prolong the agony of adjustment. Again, understanding the economics of the process is the *sine qua non* for shaping the right policy to fit the particular type and phase of inflation that is beleaguering us.

Let me return now, before closing, to several of the abiding problems of endemic inflation that are engaging the attention and efforts of economists.

An important but elusive question for the policy maker concerns the costs of inflation. Can the economist tell him anything useful and definitive on this subject? Useful, perhaps. Definitive, no. First, the economist would remind him that people continually blame inflation for crimes it does not commit. They are sure that every increase in their pay envelope is a reward for merit, every increase in prices an inflationary theft. Especially pertinent to our present shock-spiral is the observation that people "blame inflation for changes in relative prices and in real incomes that stem from market forces that have nothing to do with the course of the general price level" (Edward Foster).

Second, studies show that in a typical *U.S.* inflation, the poor have gained more in jobs and incomes than they have lost in higher prices. But in the present inflation, prices have shifted sharply against the poor, and any initial gains they may have made in jobs and income in 1973 have been more than offset by the losses incurred in the deepening 1974–75 recession induced to fight inflation.

Third, at the rates of inflation experienced prior to the 1973–75 explosion, most economists find it difficult to believe that the costs of inflation—mostly in redistributional effects, but with some distortion in resource allocation—hold a candle to the welfare losses of substantial add-ons to unemployment. Fourth, however, when inflation reaches double-digit levels, the costs in terms of the social conflicts and tensions it generates and the uncertainties and loss of confidence in the dollar yardstick it may breed are important intangibles that economists cannot ignore, yet have not been able to quantify. We need to understand far more about what unsettles and upsets people about inflation, how this affects their economic behavior, and what economic costs result. Clearly, in an economy where inflation is endemic, the balance between its gains and losses deserves intensive further study.

Another important question is this: How much of the present run-up in prices of foodstuffs, oil, and raw materials is a transitory phenomenon, how much is a one-time shift to a new plateau, and how much represents a new upward trend? Economists have trained the guns of price theory and price elasticity estimation on these questions in the case of oil and several other basic materials. They generally come up with more optimistic answers for five to ten years hence than for the near-term. But much of the answer lies in geo-political, meteorological, and similar puzzles—for example, the effectiveness of oil and other raw material cartels, the pace of world population increases and income growth, and the possibility of a dry, cold phase in world weather—that lie largely or wholly beyond the reach of economic analysis.

What we do know is this: The 1950's and the 1960's were a period of gently declining or roughly stable world prices for raw materials or foodstuffs. Now, rising population, industrialization, income, and aspirations may put such pressure on the world's supply capabilities that while we are not nearing any Club-of-Rome ultimate limits, we may for some time exceed the speed limits of stable expansion. If so we may have passed an inflection point in the price trends of basic inputs to the economy (see Walt Rostow). The mild downward trend of the 1951–71 period facilitated the rise in real incomes of urban workers side-by-side with rising profits. If this trend is reversed, rising income claims will generate greater strains, and the Phillips curve tradeoff will take place around a higher inflation constant. Economic analysis of long-run supply prices of basic commodities using alternative assumptions regarding world political, weather, and economic trends could be a useful aid to rational economic planning.

Coming back into the domain of economics as such, one should take account of the important new thinking and efforts now being devoted to the continuing mysteries of industrial pricing policies and the role of fixed-rule (generally, mark-up) pricing as a shield against uncertainty. Answering the question of how, and how fast, supply-shifts in the auction markets or market-oriented sector are transmitted through the rule-determined sector—where certain relativities seem to be maintained in the structure of prices (and wages)—is essential to an understanding of structural inflation (see Piore).

In turn, this analysis will strongly influence thinking on government intervention in private wage-price and perhaps also supply-demand decisions. If the wage-price structure is indeed fairly rigid and if supply- and demand-shifts set off an inflationary spiral, the "natural market forces" will not readily make the necessary supply-demand adjustment in any case. Wage-price restraint or controls would not be supplanting some supple and efficient resource allocation mechanism, yet would insert a circuit breaker into the inflationary spiral. This view of the world would also suggest that government action to stimulate supply and suppress demand at certain pressure points in the economy might well pass the test of economic efficiency. In pursuing these questions and hypotheses, the economist will be laying a firmer conceptual and empirical foundation for specifying the areas and circumstances in which intervention may be the lesser evil.

One should not leave the subject of economists' contributions to analysis and prescription on the inflation problem without mention of the intriguing attempt of the Brookings Panel on Economic Activity to bring the best analytical and empirical efforts of economists to bear directly on the problems and puzzles that confront the policymaker. In relation to inflation, the Panel has focussed much of its attention on such questions as the structure of labor markets, the Phillips curve relationship and wage equations, the costs of unemployment, price behavior in specific sectors like foodstuffs and oil, and the role of fiscal and monetary policies. Apart from the significant contributions that have been made to understanding these problems, and to bringing academic work into closer contact with current policy problems, the Brookings Panel is an interesting and perhaps unique exercise in "continuing confrontational econometrics." Responding to the kinds of criticisms quoted earlier in my remarks, the Brookings Panel combines rigorous quantitative testing with continuing surveillance by one's peers to assure that the investigator (a) looks beyond mathematics and makes his assumptions and relations conform to common sense, (b) spells out the implications of his econometrics and, if they are implausible, tries again, and (c) constantly keeps asking questions of the model. With the Panel now going into its sixth year of thrice-yearly meetings, previous analyses become not undisturbed museum pieces, but grist for the mill of constant retesting under the harsh light of reality and peer-group criticism.

I have dealt at some length with the substance of economists' work and findings on inflation because mere assertions of progress would hardly suffice to demonstrate what's right with economics in this most vulnerable area. The fact that there are no final or comprehensive answers has not kept economists from making significant distinctions, analyses, and measurements that equip policy makers with better means of judging the policy tradeoffs and determining how to improve the fit of policy-to-problem for the different types and stages of inflation. When policy makers fail to heed these lessons, as in 1974, both the economy and the economist feel the backlash.

Throughout this discourse, I have time and again been tempted to kick over the traces I fastened on myself and give voice to my own criticisms, dissatisfactions, and admonitions. But since an unholy (and unwitting) alliance of my colleagues and outside critics has amply and ably taken care of this, I felt it best to stay within my constraints in the interest of doing what I could do to redress the balance. As economists, we have many sins, none deadly, to confess. But these are far outweighed by the virtues, all quite lively, that we can legitimately profess.

REFERENCES

H. Aaron, "A New View of Property Tax Incidence," *Amer. Econ. Rev. Proc.*, May 1974, *54*, 212–21.

B. R. Bergmann, "Economist, Poll Thy People," *New York Times* "Points of View," Nov. 3, 1974.

W. H. Beveridge, *Full-Employment in a Free Society*, London 1944.

J. H. Blackman, "The Outlook for Economics," *Southern Econ. J.*, Apr. 1971, *37*, 385–95.

A. S. Blinder and R. M. Solow et al., "Analytical Foundation of Fiscal Policy," *Economics of Public Finance*, Washington 1974, 3–118.

K. Boulding, "Economics As a Moral Science," *Amer. Econ. Rev.*, Mar. 1969, *59*, 1–12.

H. R. Bowen, "Toward a Humanist Economics," *Nebraska J. Econ. Bus.*, Autumn 1972, *11*, 9–24.

E. Foster, "Costs and Benefits of Inflation," Fed. Res. Bank Minneapolis, Mar. 1972.

J. K. Galbraith, "Power and the Useful Economist," *Amer. Econ. Rev.*, Mar. 1973, *63*, 1–11.

F. H. Hahn, "Some Adjustment Problems," *Econometrica*, Jan. 1970, *38*, 1–17.

Robert E. Hall, "The Process of Inflation in the Labor Market," *Brookings Papers*, Washington 1974, *2*, 343–410.

Robert L. Heilbroner, "Economics as a 'Value-Free' Science," *Soc. Research*, Spring 1973, *40*, 129–43.

W. W. Heller, *New Dimensions of Political Economy*, Cambridge 1966.

H. G. Johnson, "The Economic Approach to Social Questions," *Economica*, Feb. 1968, *35*, 1–21.

———, "Scholars as Public Adversaries: The Case of Economics," in C. Frankel, ed., *Social Science Controversies and Public Policy Decisions*, ch. 12, forthcoming.

J. M. Keynes, *How to Pay for the War*, London 1940.

W. Leontief, "Theoretical Assumptions and Nonobserved Facts," *Amer. Econ. Rev.*, Mar. 1971, *61*, 1–7.

A. P. Lerner, "Functional Finance and the Federal Debt," *Soc. Research*, Feb. 1943, *10*, 38–51.

D. MacDougall, "In Praise of Economics," Presidential address to the Royal Economic Society, June 1974.

S. J. Maisel, "The Economics and Finance Literature and Decision Making," *J. Finance*, May 1974, *29*, 313–22.

R. A. Musgrave, *The Theory of Public Finance*, New York 1959.

G. Myrdal, *Asian Drama: An Inquiry into the Poverty of Nations*, Vol. I, New York 1968.

W. Nordhaus and J. Shoven, "Inflation 1973:

The Year of Infamy," *Challenge,* May-June 1974, *17,* 14–22.

A. M. Okun, *The Political Economy of Prosperity,* Washington 1969.

J. A. Pechman and B. A. Okner, *Who Bears the Tax Burden?,* Washington 1974.

G. L. Perry, "The Petroleum Crisis and the U.S. Economy," prepared for the Brookings Conference on the Impact of Higher Oil Prices on the World Economy, Nov. 1974 (to be published).

E. H. Phelps Brown, "The Underdevelopment of Economics," *Econ. J.,* Mar. 1972, *82,* 1–10.

M. Piore, "Curing Inflation with Unemployment, Outmoded Notions of Supply and Demand," *New Republic,* Nov. 2, 1974, 27–31.

A. M. Rivlin, "Why Can't We Get Things Done," *Brookings Bull.,* Spring 1972, *9,* 5–9.

———, "Social Experiments: The Promise and the Problem," *Evaluation,* 1973, *1,* 77–78.

———, *Systematic Thinking for Social Action,* Washington 1971.

M. J. Roberts, "On the Nature and Condition of Social Science," *Daedalus,* Summer 1974, *103,* 47–64.

W. W. Rostow, "Political Economy in a Time of Scarcity: How to Get From Here to There," *Naval War College Rev.,* Sept.-Oct. 1974, *27,* 32–45.

N. Ruggles, *Economics,* Englewood Cliffs 1970.

C. L. Schultze, "Is Economics Obsolete? No, Underemployed," *Saturday Review,* Jan. 22, 1972.

———, *Recent Inflation in the U.S.,* Joint Economic Committee, work. pap. no. 1, Congress of the United States, Washington 1959.

———, "The Reviewers Reviewed," *Amer. Econ. Rev. Proc.,* May 1971, *61,* 45–52.

T. W. Schultz et al., "Human Capital: Policy Issues and Research Opportunities," in his *Human Resources,* Vol. VI, New York 1972.

C. S. Shoup et al., *Public Expenditures and Taxation,* New York 1972.

R. M. Solow, "Science and Ideology in Economics," *Publ. Interest,* Fall 1970, *21,* 94–107.

———, "The State of Economics—Discussion," *Amer. Econ. Rev. Proc.,* May 1971, *61,* 63–68.

———, "What the Intelligent Citizen Should Know About Inflation," *Publ. Interest,* forthcoming 1975.

G. J. Stigler, "The Economist and the State," *Amer. Econ. Rev.,* Mar. 1965, *55,* 1–18.

P. Sweezy, "Capitalism, for Worse," *Monthly Rev.,* Feb. 1974, *25,* 1–7.

J. Tobin, *The New Economics One Decade Older,* Princeton 1974.

———, "Cambridge (U.K.) vs. Cambridge (Mass.)," *Publ. Interest,* Spring 1973, *31,* 102–09.

F. A. von Hayek, "Inflation and Unemployment," *New York Times* "Points of View," Nov. 15, 1974.

B. Ward, *What's Wrong With Economics?* New York 1972, p. 12.

G. D. N. Worswick, "Is Progress in Economic Science Possible?" *Econ. J.,* Mar. 1972, *82,* 73–79.

Supply and Salary Adjustments to the Changing Science Manpower Market: Physics, 1948-1973

By R. B. FREEMAN*

The collapse of the science manpower market in the late 1960's raises important questions about the way in which high level labor markets adjust to sharp changes in economic conditions. What happens to salaries when demand for specialists declines greatly? How does the supply of new entrants respond to market ups and downs? What are the implications of the adjustment process for future manpower developments and current methods of manpower forecasting?

This paper examines these questions with respect to physics, which, as one of the first specialties to undergo the turnaround from relative shortage to surplus, offers an especially fruitful case for study. It finds real salaries and the supply of new entrants to physics to be highly sensitive to market conditions. Economic responsiveness critically affects the likely future state of the field and should be taken into account in the standard forecast methodology (see Alan Cartter). Section I outlines the principal post-World War II changes in the physics manpower market, the phenomenon under investigation. Section II analyzes the supply behavior of students and the determination of physics salaries in the framework of a recursive market model (see the author, 1971). The estimated model is used to forecast the future state of physics manpower, yielding strikingly different results from those obtained by standard techniques (see Cartter). These differences highlight the sensitivity of the labor market as an allocative device and the critical role of supply and demand adjustments in manpower analyses.

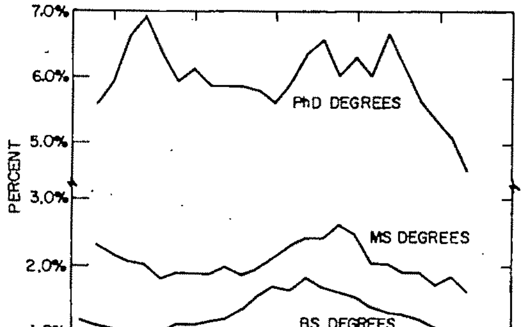
I. The Physics Job Market

The 1960's turnaround in the state of the physics labor market was marked by three important developments. First, as depicted in Figure 1, there was a substantial change in the flow of students into the field, with large and increasing numbers choosing physics in the late 1950's-early 1960's and fewer entering in the late 1960's. Panel A, which records the number of physics degrees relative to total male degrees from 1948 to 1972, shows, for example, that the relative number of B.S. degrees in physics rose in the 1950's to a peak of 1.84 percent in 1962 and then dropped nearly in half to 0.96 percent in 1972. Ph.D. degrees in the field plummetted downward from 6.7 to 4.6 percent of the total in the late 1960's (1967–72) and declined absolutely at the onset of the 1970's (−7 percent from 1970 to 1972). M.S. degrees followed a pattern similar to that of B.S.s and Ph.D.s, peaking in 1964, just two years after the B.S. peak and three years before the Ph.D. turning point.

The ratios of undergraduate enrollments

---

* Harvard University. The research underlying this report was supported by a National Science Foundation grant to the Center for Policy Alternatives at the Massachusetts Institute of Technology. Any opinions, findings, conclusions, or recommendations expressed herein are those of the author and do not necessarily reflect the views of the Center for Policy Alternatives, the Massachusetts Institute of Technology, or the National Science Foundation. F. Amacher and S. Sonnabend proved useful research assistants.



Degrees, 1948-71, U.S. Bureau of Census (1949-1972); 1972 estimated using American Institute of Physics Pub. R-151.10 (1973).
Enrollments, American Institute of Physics, Pub. R-151.10 (1973).

FIGURE 1

in physics to the population of prospective enrollees in panel B reveal even more striking changes in the supply of students between the 1950's and 1960's. At the undergraduate level, the proportion of male freshmen selecting physics in their junior year rises from 1.24 percent in 1953 to 2.31 percent in 1960 and then drops to a postwar low of 0.99 percent in 1972. Despite a near doubling in the number of undergraduate males, junior and senior year physics majors decline from 14,900 (1961) to 11,400 (1972). Graduate enrollments, having increased from 3.00 percent of all male graduate students in 1955 to 4.14 percent in 1961, trend downward sharply to 1.73 percent in 1972 with a 24 percent drop in absolute numbers at the end of the 1960's. First-year graduate enrollments, an especially sensitive measure of the adjustment margin of physics supply, fall by a third from 1961 to 1972 and decline sharply even when compared to the decreasing number of B.S. graduates. In short, there were large changes in the supply of students to physics in the postwar period, with the upward trend of the 1950's arrested and reversed in the 1960's.

The second major development was a significant fall in real salaries during the period of market downturn, after a decade or so of substantial increases in salaries. This sharp change in the economic status of physicists is examined in Table 1, which records percentage changes in physics and other salaries or earnings for various subperiods since World War II. While the patterns of change vary somewhat depending on the source of the data or the level of training, a reasonably clear picture emerges of physics salary developments.[1] During the 1954–62 period encompassing Sputnik (October 1957) and the Research and Development (*R&D*) boom, the salaries of Ph.D. and M.S. physicists, who constitute the bulk of the profession (80 percent), and the starting salaries of B.S. physicists (but not the annual earnings reported by the National Science Foundation (*NSF*)) increased more rapidly than those of comparison groups. After 1962, by contrast, the rate of change of physics salaries fell increasingly behind that in male professional earnings, annual earnings of full-time employees, and average hourly earnings. Between 1962 and 1968, physics Ph.D. earnings increased by 22 percent, M.S. earnings by 26 percent, and B.S. earnings by 25 percent compared to a 36 percent gain for male professionals. At the end of the 1960's and the onset of the 1970's, the income position of physicists declined more rapidly, with real as well as relative salaries falling. American Institute of Physics (*AIP*) surveys show, for example, no gain in starting Ph.D. and M.S. salaries and a 5.8 percent decline in B.S. salaries in the 1970–72 period when consumer prices advanced by 12 percent; College Placement Council (*CPC*) data show rough stability in money salaries over the same period and into the 1973 recruitment season. The net effect of this decline is to reduce relative physics earnings to pre-Sputnik levels or lower: in 1954 the premium of Ph.D. physics earnings over male professional earnings was 58 percent; in 1962, 89 percent; by 1970, it had fallen back to 57 percent. From 1970 to 1972 the ratio of *AIP* starting rates to professional earnings records a further drop of 12 percentage points (1.25 to 1.13). It is apparent from these changes that relative salaries in physics have been highly sensitive to market conditions and roughly as responsive to market declines as to booms.

---

[1] The major difference by level of training is the rapid increase of M.S. and B.S. salaries between 1951 and 1954 compared to the slow increase in Ph.D. salaries. The major difference by source of data is between B.S. salaries reported by the National Register, which increases only moderately from 1951 to 1962, and those of starting physicists reported by placement offices, which increase rapidly in that period.

Third, there was a sizeable shift in federal expenditures on basic research and development, atomic energy, and graduate training which had an exceptional impact

TABLE 1—CHANGES IN PHYSICS SALARIES, 1948–73
(percentage change in current dollar salaries)

	1951–54	1954–62	1962–68	1968–70	1970–72	1972–73
Ph.D.						
Earnings	9.2	67.7	22.3	8.8	—	—
Starting Salaries	—	54.4[a]	18.1	6.6	3.3	1.3
Starting Salaries	—	—	—	12.0	−0.1	—
M.S.						
Earnings	21.4	55.4	26.0	11.9	—	—
Starting Salaries	—	—	30.5	11.8	−0.3	2.8
Starting Salaries	—	—	—	5.8	0.0	—
B.S.						
Earnings	28.6	37.2	25.0	15.9	—	—
Starting Salaries	—	50.9[a]	30.6	10.1	−1.1	3.3
Starting Salaries	—	—	—	10.0	−5.6	—
Comparison Groups						
Male Professionals	20.5	40.1	36.4	17.4	9.9[b]	—
Annual Earnings	14.0	38.1	31.4	13.8	13.2[b]	—
Average Hourly Earnings	13.8	34.5	28.4	13.0	13.3	5.8[c]
CPI	3.4	12.6	15.0	11.6	11.9	—

Sources: Earnings from *NSF American Science Manpower*, various editions, except 1951, from U.S. Federal Security Agency (1952). Starting Salaries from *CPC*, 1960–73, except for 1954–62, as noted. Starting Salaries from *AIP*, 1968–72a,b. Professional Earnings, U.S. Bureau of Census, 1948–72. Annual Earnings, U.S. Department of Commerce, 1966–72. Average Hourly Earnings and *CPI*, U.S. Bureau of Census, 1948–73.

[a] Calculated from salaries of Massachusetts Institute of Technology graduates, with data obtained by the Placement Office at M.I.T. The Ph.D. salaries from 1955 to 1961; B.S. salaries from 1955 to 1962.

[b] Extrapolation for 1971–72 at rate of increase of average hourly earnings.

[c] Estimated as change from April 1972 to April 1973.

TABLE 2—FEDERAL SUPPORT OF PHYSICS, 1951–72
(in constant 1957–59 millions of dollars)

	1951	1954	1962	1968	1970	1972
Demand						
1) Federal *R&D* Spending as percent	$1,638	3,353	8,789	12,370	10,868	10,064
GNP	0.45	0.86	1.65	1.73	1.51	1.32
2) Atomic Energy Commission Spending	$ 547	1,110	2,558	2,069	1,851	1,702
as percent *GNP*	0.15	0.28	0.48	0.29	0.26	0.24
3) Federal Physics Research	$ —	137[a]	353	511	398	377
per Physicist[c]	$ —	7	14	16	11	10
4) Federal Basic Physical Science Research	$ —	88[a]	573	876	781	809
per Physicist[c]	$ —	7	23	27	22	22
5) Percent Physicists with Federal Support	—	—	66	62	55	—
Supply						
6) Number of Physical Science Fellowships	—	406	1,684	5,960	4,459[b]	2,964[b]
as percent Graduate Enrollment	—	3	11	21	16	6
7) Average Value of Fellowship	$ —	1,688	3,767	3,856	3,420[b]	3,170[b]
as percent Professional Salaries	—	32	58	50	42	37

Sources: 1) *NSF*, 1972b; 2) U.S. Atomic Energy Commission, 1959–72; 3), 4) U.S. Atomic Energy Commission, with number of physicists from *American Science Manpower*, various editions; 5) *American Science Manpower*, various editions; 6), 7) *Federal Interagency Committee on Education*, 1970, 1972.

[a] Used 1959, earliest year with data.

[b] Estimated from number or value of all federal fellowships, assuming constant ratio of physical sciences to total.

[c] In thousands.

on physics and other scientific specialties. This shift appears to be the chief cause of the downturn in the science market (see the author, 1972b). Table 2 summarizes data pertaining to federal spending and activities that have impinged on physics in the 1950's and 1960's. Line 1 shows the postwar fluctuation of total *R&D*: the 1950's spurt in expenditures, the 1960's levelling of the rate of increase, and the 1968–72 decline in real and relative spending. A similar pattern is displayed in lines 2–5, which deal with activities more directly linked to physics: federally funded physics research (line 3), physical science basic research (line 4), percentage of physicists with federal support (line 5), and Atomic Energy Commission spending, which turns down in 1962 (line 2). Federal fellowship aid (lines 6–7) underwent a different path of increase and decrease, with the number of awards growing rapidly from 1962 to 1968 and declining thereafter. Because federal fellowships declined after the expenditure plateau, students supported in the 1960's graduated during the spending decline of the early 1970's, contributing to the manpower crisis of that time.

The turnaround in the market for physicists and other scientific specialists described above and the impending "demographic crunch" on higher education have spawned a crop of manpower requirements forecasts of the future (see Cartter, *NSF* (1971), Dael Wolfe and Charles Kidd) based on extrapolations of current and past developments. While useful in directing attention to possible supply/demand imbalances, the forecasts can be criticized for ignoring the effects of salaries and economic behavior on market developments,[2] possibly leading to erroneous or exaggerated statements of problems (see

---

[2] It should be stressed that the well-known Cartter forecast depends on a decline in the number of persons of college age in the 1980's, a relatively hard piece of evidence on which to base forecasts.

Glen Cain, the author, and Lee Hansen; the author and David Breneman). The degree to which forecasts are potentially misleading depends, all else the same, on the way in which the science manpower market adjusts to changes in economic conditions. How responsive is the market mechanism? Are supply and demand adjustments to economic incentives first- or second-order effects in manpower forecasts?

II. Market Adjustments and Forecasting in a Recursive Model

To answer these questions in the physics case, we analyze the observed pattern of change in enrollments, degrees, and salaries with a simple recursive model of the labor market and use the model to predict the number of Ph.D.s in the field in the 1970–80 decade. Variants of the model, which exploit the time delay in the production of graduates due to lengthy training periods to identify supply and demand adjustment processes, have been applied to other high level occupations with some success (see the author, 1971, 1972a,b, and forthcoming). In simplest form, focusing on starting graduates, the model contains a supply schedule linking the number of students entering a field to starting salaries and related incentives; a demand equation with the number demanded dependent on salaries and exogenous demand-shift variables;[3] and market clearing. Because of the lag in training specialists (upwards of 4–5 years for Ph.D. physicists), the number of new entrants depends on past market conditions and is predetermined with respect to the demand and market clearing equations. This makes salaries rather than quantities the appropriate dependent variable in the demand equation and creates a recursive "cobweb" structure that generates endogenous cyclical fluctuations. The recursive model differs greatly from the Arrow-Capron model of scientific markets wherein salaries, rather than supplies, react slowly to changes in the market. When, as seems reasonable, the disturbances in the supply and salary equations are independent, the equations are appropriately estimated by least squares.

Depending on data availability and the problems of concern, the model can be amplified to focus on the complexities of expectations formation (see Marc Nerlove, John Muth), adjustments in salaries due to bargaining during recruitment, decisions to complete as well as initiate training, interactions between the salaries and demand for younger and older specialists, and so forth. The current paper concentrates on the basic aspects of market adjustment: the responsiveness of students to market incentives, the factors determining starting salaries, and the "cobweb supply relation" which links supply to past levels of supply and demand. The cobweb relation, obtained by substituting the salary equation into the basic supply equation,[4] highlights the endogenous fluctuations that characterize the market.

A. *Supply Behavior*

The decision to study physics is examined in Tables 3 and 4 which relate (in *log* form) enrollments and degrees to salaries and a measure of alternative opportunities (male professional earnings). The salary variables are dated at the June preceding the date of enrollments used in the enrollment regressions and at the June 3,

---

[3] Selection of exogenous demand variables is difficult in analysis of factor demands. In the case of physics, federal *R&D* policies are taken as the major exogenous factor at work, making *R&D* expenditures exogenous.

[4] Formally, with the simple model

(1) $S = aW(-1)$ [Supply equation]

(2) $W = bS + cZ$ [Wage equation]

where S = supply, W = wages, Z = level of demand. Substituting we obtain

(3) $S = abS(-1) + acZ(-1)$ [Cobweb equation]

TABLE 3—REGRESSION ESTIMATES OF THE PROPORTION OF STUDENTS IN PHYSICS, 1954–72[a]

Dependent Variable and Period Covered	Constant	Physics[b] Salary	Professional Earnings	Lagged Proportion	R^2	S.E.E.
1) Junior Majors/Male Freshman Lagged 2 years 1959–72	3.74	3.64	−3.50		.76	.12
		(0.70)	(.53)			
2) Junior Majors/Male Freshman Lagged 2 years 1959–72	4.24	1.28	−1.54	.69	.93	.06
		(.56)	(.45)	(.12)		
3) Graduate Enrollments 1954, 1957–72[c]	10.02	2.23	−3.24		.86	.10
		(.37)	(.42)			
4) Graduate Enrollments 1954, 1957–72[c]	4.55	0.61	−1.04	.87	.97	.05
		(.36)	(.47)	(.16)		
5) First-year Graduate Enrollments/B.S. Physics Graduates 1954, 1960–72[d]	5.66	1.05	−1.49		.72	.08
		(.27)	(.31)			
6) First-year Graduate Enrollments/B.S. Physics Graduates 1954, 1960–72[d]	4.23	.42	− .82	.73	.90	.05
		(.37)	(.28)	(.22)		
7) B.S. Degrees, 1957–72	8.06	5.53	−5.36		.84	.09
		(.90)	(.73)			
8) B.S. Degrees, 1957–72	6.20	2.16	−2.41	.68	.97	.04
		(.61)	(.52)	(.09)		
9) M.S. Degrees, 1956, 1959–72[e]	5.75	1.90	−2.49		.70	.09
		(.41)	(.52)			
10) M.S. Degrees, 1956, 1959–72[e]	3.40	1.00	−1.31	.59	.84	.07
		(.44)	(.57)	(.21)		
11) Ph.D. Degrees, 1956, 1959–72[e]	3.93	1.22	−1.71		.71	.08
		(.34)	(.39)			
12) Ph.D. Degrees, 1956, 1959–72[e]	2.75	.79	−1.03	.68	.81	.06
		(.35)	(.46)	(.32)		

Source: See Tables 1, 2, and Figure 1, with salaries calculated by linking series as described in fn. 6.

[a] The dependent variables are the logarithm of the proportion of students at the same level of training, unless otherwise specified. Independent variables are also in *logs*. Numbers in parentheses are standard errors of estimate.

[b] B.S. salaries used in line 1, 2, 7, 8; Ph.D. salaries in lines 3–6, 11, 12; M.S. salaries in lines 9–10; all salaries deflated by *CPI*. Salaries lagged 3 years in lines 7–8, 2 years in lines 8–9, 5 years in lines 11–12.

[c] Omitted 1958, 1959, 1961, and 1963 due to absence of physics salaries for the period of the supply decision.

[d] Omitted 1961, 1963 due to absence of data.

[e] Omitted 1960, 1961, 1963 due to absence of salary data for period of supply decision.

2, and 5 years earlier in the B.S., M.S., and Ph.D. regressions, respectively. Enrollments and degrees provide alternative, related measures of the same underlying supply behavior, with the former reflecting the immediate, and the latter, the ultimate impact of market incentives on the number of new entrants. The link between initial enrollments and degrees which results from the decision to complete a program (see the author, 1971) is not treated here due to the limited number of data points. The dependent variable in Table 3 is the proportion of students in physics; the dependent variable in Table 4 is the absolute number of students, yielding somewhat different estimating equations.[5] The starting salary variables are viewed as proxies for anticipated lifetime incomes, appropriately discounted. Lagged dependent variables are used to estimate the time delay in supply responses in the framework of a simple lagged adjustment model. Because we lack a continuous series of salaries in physics

---

[5] The equations differ because regressions with numbers as the dependent variables do not necessarily support the homogeneity postulate of the proportion regressions. Indeed, when the lagged number of physics students is entered in the regressions, the coefficient for all students becomes insignificant, and is dropped in the results shown in Table 4. As the homogeneity postulate remains a priori plausible, it was deemed fruitful to present the two sets of regressions.

TABLE 4—REGRESSION ESTIMATES OF PHYSICS ENROLLMENTS AND DEGREES 1954–72[a]

Equation	Dependent Variable	Constant	Physics Salaries	Professional Earnings	Lagged Number	R^2	S.E.E.
1)	Junior Majors 1957–72	3.46	1.15 (.50)	−.90 (.28)	.71 (.11)	.94	.06
2)	Graduate Enrollments 1954, 1957–72[b]	0.70	1.02 (.24)	−.49 (.23)	.65 (.16)	.99	.03
3)	First-Year Graduate Students, 1954, 1960–72[b]	8.06	.87 (.60)	−1.04 (.30)	.42 (.38)	.81	.06
4)	B.S. Degrees, 1957–72	1.72	2.29 (.88)	−1.26 (.50)	.40 (.17)	.96	.04
5)	M.S. Degrees, 1956, 1959–72[c]	−1.10	.93 (.24)	−.27 (.36)	.67 (.10)	.99	.05
6)	Ph.D. Degrees, 1956, 1959–72[c]	−.68	.82 (.25)	−.42 (.38)	.83 (.11)	.99	.03

Source: See Tables 1, 2, and Figure 1, with salaries calculated by linking series as described in fn. 6.

[a] Dependent variables are the number of students at the level specified. All variables are in *log* form. Numbers in parentheses are standard errors of estimate.

[b] B.S. salaries in lines 1, 4; Ph.D. salaries in lines 2, 3, 6; M.S. salaries in line 5; all salaries deflated by the consumer price index. Salaries lagged 3 years in line 4, 2 years in line 5; 5 years in line 6.

[c] Omitted 1960, 1961, 1963, due to absence of salary data for period of supply decision.

over an extended time period, the calculations are based on a composite series obtained by linking overlapping *NSF*, *CPC*, and related data by least squares estimation.[6]

The computations of Table 3 reveal an extremely close relation between salary incentives and the number of physics students, suggesting that the ups-and-downs in supply shown in Figure 1 result from economically responsive behavior. Physics salaries obtain a sizeable positive coefficient in all of the regressions, and professional earnings a negative coefficient, often of similar magnitude. The salary variables account, by themselves, for three-fourths or more of the variation in the physics share of enrollments and degrees (odd-numbered regressions), with elasticities ranging from approximately unity (line 5) to over five (line 7). With the addition of the lagged endogenous variable, salary coefficients decline substantially (even-numbered regressions), indicative of different short- and long-run supply elasticities. Larger undergraduate than graduate elasticities presumably reflect the greater flexibility of persons not yet committed to science careers. The only divergent result in the table occurred in the Ph.D. regressions where an initial calculation relating the physics share to salaries lagged five years (consistent with the other lag patterns) yielded weaker results than that given in the table with the two-year lag,[7] suggesting substantial economic responsiveness in the decision to complete the doctorate program. As the physics doctorate typically takes four or five years to complete, the effect of salaries on the initial enrollment decision and on eventual graduation is likely to involve complex lag structures, on which our limited number of observations provides little light.

---

[6] Milton Friedman and Gregory Chow and An-loh Lin describe the optimal methods of extrapolation, interpolation, and linking series. In our case, the Ph.D. and M.S. series are obtained by regressing *CPC* (1960–73) data on *NSF* data (1951, 1954, 1957, 1960, 1962, 1964, 1966, 1968, 1970) and using the predicted salaries for the pre-1960 years in which *NSF* figures exist. The B.S. salary series is obtained by regression of *CPC* figures on a series obtained from the M.I.T. Placement Office for graduates of that school. All of the data are available on request.

[7] The regression with the five-year lag yielded a coefficient for physics salaries of 0.73, with a standard error of 0.48 and one for professional salaries of −1.42, with a standard error of 0.63.

Additional support for the economic interpretation of the changing supply of students to physics is given in Table 4, which focuses on the absolute rather than relative number in the field. Salaries are the key variables in these calculations, with the short-run elasticity of supply to physics salaries estimated at near unity for graduate students and over two for B.S. recipients. Interpreting the lagged dependent coefficients as adjustment parameters, the long-run elasticities of degrees to physics salaries range from 2.8 (M.S.) to 4.6 (Ph.D.) and 4.8 (B.S.). The summary statistics, high R^2, small standard errors of estimate, show that the simple supply model does a surprisingly good job in tracking the market turnaround, which suggests its value in forecasting.

B. *Provisos*

Although the calculations in Tables 3 and 4 support our analysis, they fall short of a "complete" explanation of supply developments in several ways. First, fellowship availability was omitted from the calculations due to the absence of data, with a potential bias on the estimated salary coefficients. The direction of the bias, which depends on the correlation of fellowships with supply (presumed positive) and of fellowships with salaries is unclear,[8] for the latter would appear to have been positive in the 1950's and early 1960's but negative during the period of greatest increase in fellowship support (1962–68). Second, the quality of physics students and programs has also been ignored, again with unclear consequences. In this case, while the less qualified are more likely to provide the adjustment margin in supply, high quality institutions (heavily dependent on federal *R&D* programs) have reduced enrollment more than low quality colleges and universities (see *NSF*, 1972a). Third, the simple lagged adjustment model is at best a crude first-order approximation to the underlying process of forming salary expectations and adjusting to new market conditions. Various alternative models allowing, say, for different lags with respect to physics salaries than to the broader professional earnings, have not been explored due to the lack of a complete salary series for the postwar period. Fourth, professional earnings have been used as a measure of the alternative opportunities facing students at various decision points, although in principle alternatives will vary by level of training.[9] These provisos notwithstanding, the evidence indicates significant student supply responsiveness to the changing science manpower market.

C. *Cobweb Supply Patterns*

An alternative perspective on supply behavior is given by the cobweb supply equation which links the career decision directly to the level of demand and predetermined numbers of graduates. The cobweb equation can be obtained, as noted earlier, by substituting the salary determination relation for salaries in the supply equation of the recursive model. It may also be obtained from a "job opportunities" model of career decisions in which opportunities depend on the balance between levels of demand and persons seeking work. Information about opportunities as well as salaries is likely to be useful in evaluating the current state of the market and in forecasting future possibilities. Deletion of salaries from the supply equation permits estimation in the absence of "good" salary data, which is especially fruitful in the case at hand.

---

[8] As Zvi Griliches shows, the omitted variable bias depends on the product of the least squares coefficients of the regression of the dependent variable on the excluded variable and of the regression of the included explanatory variable on the excluded variable.

[9] In theory, the alternative income affecting decisions will change each year as cumulated human capital increases. One experiment designed to provide a better measure of alternatives (which is possible with available data) is to treat B.S. salaries rather than professional earnings as the relevant alternatives to graduate students. This is, however, by no means optimal since many undergraduate majors do not view B.S. physics work as the proper alternative. Calculations with B.S. salaries support the results in the text. The regression of first-year enrollments divided by B.S. graduates on Ph.D. and B.S. salaries, for example, yields coefficients of 1.21 and −1.69, with standard errors of .93 and 1.23. The regression of the physics share of M.S. degrees on M.S. and B.S. salaries two years earlier yields 2.12 and −3.92 for the estimated coefficients, with standard errors of 1.04 and 1.32.

Table 5 contains estimates of the cobweb supply relation in physics for the postwar period. In these computations, enrollments or degrees are regressed on the lagged dependent variable, professional earnings, *R&D* spending (indexing the level of demand), and degrees in the year of the supply decision. In the degree regressions, the explanatory variables are lagged three (B.S., lines 1 and 2), four (M.S., lines 3 and 4), and five (Ph.D., lines 5 and 6)

TABLE 5—COBWEB SUPPLY EQUATION[a] 1950–72

Equation	Dependent Variable, Time Coverage and Lag	Constant	Graduates[b] (0)	Graduates[b] (−1)	*R&D*	Professional Earnings	Lagged Number	R^2	*S.E.E.*	Adjusted *D.W.*
1)	B.S. Degrees, 1950–72	7.20	−.42		.63	−.87	.76	.989	.043	1.77
			(.05)		(.11)	(.26)	(.09)			
2)	B.S. Degrees, 1950–72	8.24	−.17	−.24	.78	−.97	.55	.992	.040	0.65
			(.14)	(.12)	(.12)	(.25)	(.13)			
3)	M.S. Degrees, 1950–72	5.40	−.13		.64	−.88	.61	.992	.046	4.21
			(.05)		(.11)	(.33)	(.08)			
4)[d]	M.S. Degrees, 1951–72	8.15	−.12		.64	−.81	.60	.999	.035	—
			(.03)		(.06)	(.19)	(.05)			
5)	Ph.D. Degrees, 1950–72	2.93	−.14		.45	−.41	.72	.988	.055	1.10
			(.08)		(.12)	(.55)	(.08)			
6)[d]	Ph.D. Degrees, 1951–72	1.81	−.15		.40	−.27	.74	.995	.055	—
			(.09)		(.12)	(.59)	(.06)			
7)	Junior Majors, 1955–72	6.16	−.37		.44	−.69	.81	.961	.061	—
			(.12)		(.24)	(.42)	(.16)			
8)[c]	Junior Majors, 1956–72	7.53	−.30		.40	−.63	.83	.965	.064	—
			(.12)		(.21)	(.37)	(.14)			
9)	First-Year Graduate Students, 1954, 1959–72	7.08	−.42		1.10	−.76	—[e]	.967	.047	—[e]
			(.12)		(.11)	(.66)				
10)	Graduate Students, 1955–72	.94	−.22[b]		.26	−.08	.87	.990	.044	2.38
			(.12)		(.19)	(.62)	(.14)			
11)[d]	Graduate Students, 1956–72	3.19	−.18		.27	−.27	.91	.998	.030	—
			(.06)		(.08)	(.30)	(.06)			

Source: See Tables 1, 2, and Figure 1, with salaries calculated by linking series as described in fn. 8.

[a] The dependent variable is the *log* of the number of degrees or enrollees, as specified. Independent variables are lagged by 3 years in lines 1 and 2; 4 years in lines 3 and 4; 5 years in lines 5 and 6; 1 year in lines 7–11. Numbers in parentheses are standard errors of estimate; all variables are in *log* form.

[b] B.S. graduates used in lines 1, 2, 7, 8; Ph.D. graduates in lines 5, 6, 9–11; and M.S. graduates in lines 3–4. Graduates with zero (0) lag are lagged the same number of years as other variables; graduates with (−1) lag are lagged an additional year.

[c] Adjusted *D.W.* statistic calculated by Durbin procedure which takes account of lagged dependent variable. The statistic is distributed, asymptotically, as the standard normal.

[d] Non-linear least squares estimates, with autocorrelation coefficients of −.66 (line 4), −.41 (line 6), −.24 (line 8), and −.76 (line 11).

[e] Omitted as insignificant; Standard *D.W.* statistic of 1.96 showed no serial correlation.

years, with graduates sometimes lagged an additional year. Precise determination of the cobweb cycles caused by the negative impact of degrees on enrollments is not possible due to collinearity between degrees one year apart. In some regressions, the table shows "best" results for graduates having the same lag as other variables (lines 3, 4, 10, 11); in other cases, an additional year lag does better (lines 5–9), while for B.S. degrees (lines 1 and 2), the cobweb effect is shared between the two. As evinced in lines 1 and 2, however, the estimated size of the effect is roughly the same regardless of the specific lag (adding the additional degree term in line 2 reduces the coefficient in line 1 from $-.42$ to $-.17$, increases its standard error, but gives the same total effect $(-.41=-.17-.24)$). Similar results are obtained for the other independent variables. More detailed investigation of autoregressive degree patterns confirms the basic results of the table (see the author, 1973b). Finally, the problem of negative serial correlation of residuals shown by the Durbin-Watson statistics adjusted to take account of the lagged dependent variable (see James Durbin) is treated in the non-linear least squares regressions of lines 4, 6, 8, and 11 where a one-period autocorrelation coefficient is estimated simultaneously with other coefficients, with little effect on the parameters of interest.

There are two important findings in Table 5. First, the key coefficient on graduates which must be negative by the model obtains the predicted sign in all regressions with especially large effects at the baccalaureate level (lines 1, 2, 7, 8) and for first-year graduate enrollments (line 9). The other explanatory variables are also accorded correctly signed and quantitatively plausible impacts on supply. Second, the regressions as a whole show that the cobweb relation provides about as good an explanation of degrees or enrollments as the supply regressions of Tables 3 and 4, with similar R^2 and standard errors of estimate.

D. *Determination of Physics Starting Salaries*

The demand or salary side of the market is investigated in Table 6 which displays regressions of physics salaries on the level of demand measured by *R&D* and numbers of graduates lagged one and two years. These regressions lend support to the recursive model with *R&D* having a substantial positive effect, and degrees a substantial negative effect on all three salary variables. As shown, *R&D* is accorded a larger effect on Ph.D. and M.S. than on B.S. salaries, presumably as a result of the greater concentration of *R&D* activity among higher degree recipients. The estimated coefficients have relatively small standard errors in regressions 1, 3, and 5 where only a single demand or supply variable is entered; addition of collinear variables in regressions 2, 4, and 6 reduces the coefficients and increases their standard errors, but (as in Table 5) preserves the sum. Hence the size but not the precise timing of the impact of *R&D* or degree on salaries is reasonably well determined: the elasticity of salaries to *R&D* ranges from .34(B.S.) to .54(M.S.) and the flexibility of salaries from minus .15 to .25.[10]

The regressions also provide information enabling us to evaluate the relative importance of demand and supply forces in the changes in physics salaries shown in Table 1. As the variance of *R&D* and degrees about their means turns out to be nearly the same in the period covered, the

---

[10] Note that the regression relates starting salaries to new entrants, not average salaries to total employment, and thus does not yield elasticities with respect to the stock of physicists. With ratios of graduates to employment of about 1 to 10, the flexibility coefficients for the entire stock would be ten times those in the text, which in turn implies that demand for all physicists is *inelastic*.

TABLE 6—SALARY DETERMINATION EQUATIONS[a] 1951–72

Equation	Dependent Variable and Time Coverage	Constant	*R&D* (−1)	*R&D* (−2)	Degrees (−1)	Degrees (−2)	R^2	*S.E.E.*
1)	B.S. Salaries,	4.38	—	.34	−.16		.95	.047
	1955–72			(.05)	(.06)			
2)		4.15	.22	.14	−.02	−.13	.96	.021
			(.11)	(.11)	(.08)	(.08)		
3)	M.S. Salaries,	3.19		.54		−.26	.98	.030
	1951–72[b]			(.03)		(.04)		
4)		2.97	.20	.36	−.12	−.13	.98	.029
			(.17)	(.19)	(.10)	(.09)		
5)	Ph.D. Salaries,	3.57		.44		−.14	.96	.037
	1951–72[b]			(.04)		(.04)		
6)		3.36	.19	.14	+.05	−.16	.97	.037
			(.26)	(.24)	(.16)	(.12)		

Source: See Tables 1, 2, and Figure 1, with salaries calculated by linking series as described in fn. 6.

[a] The dependent variable is starting salary of the specified group; salaries and *R&D* are in 1957–59 constant dollars. Dependent variables are lagged 1 or 2 years as specified (−1), (−2). All variables are in *log* form. Numbers in parentheses are standard errors of estimate.

[b] The years 1952, 1953, 1955, 1956, 1958, 1959, 1961, 1963 are omitted due to absence of salary data.

larger *R&D* parameters in the regressions implies that changed demand, due primarily to changes in federal *R&D* spending,[11] was the *chief source* of changes in physics salaries, though movements along the demand schedule due to supply developments are also significant.

E. *Market Adjustments and Forecasting*

To what extent do the preceding findings, which depict a highly responsive market adjustment process, lead to forecasts of future developments and problems that differ from those obtained by standard projection techniques? Table 7 contains predictions from the recursive model and other studies showing a marked divergence in forecasts, with strikingly fewer physics Ph.D.s anticipated for the future in our model than in even the lowest standard projection. More precisely, the table uses two versions of the estimated model to generate forecasts. Variant A employs the supply of Ph.D. degrees equation from Table 4 (line 6) and the Ph.D. salary equation from Table 6 (line 5) to generate forecasts. Since the degrees lag salaries by 4 years, predicted salaries are not needed until the 1977 degree forecast. Variant B replaces the degree equation with an equation determining first year enrollments (Table 4, line 3), assuming a similar percentage increase in Ph.D.s as in initial enrollments. In both models, *R&D* is assumed to increase at the postwar rate of growth of real *GNP* (3.7 percent) so that the *R&D* to *GNP* ratio is fixed; professional earnings are postulated to increase at their postwar trend level of 2.8 percent per year.

The important finding is that regardless of the precise equations used, our model predicts sizeable declines in physics doctorates while the other analysts foresee large increases. According to the recursive model, reductions in the supply of new

---

[11] Evidence regarding the effect of federal *R&D* on total *R&D*, which shows much greater variance in the former than the latter, is given in the author (1972b).

TABLE 7—ALTERNATIVE FORECASTS OF PHYSICS PH.D.S 1970–80

Years	Office of Education	Cartter	*NAS-NRC*	Model A	Model B
1969–70[a]	1500	1500	1500	1500	1500
1974–75	2153	1853	2153	1134	1189
1979–80	2492	2654	3350	731	786

Source: Office of Education, Cartter, *NAS-NRC* from Cartter; Models A and B, from the recursive model of Tables 4 and 5, as described in the text.

[a] The various projections use different base year numbers of Ph.D.s due to differences in the source of data. National Academy of Science-National Research Council (*NAS-NRC*) data differ somewhat from those of the Office of Education. I have set the base year number as 1500 and made projections on the basis of implicit percentage increases in the actual forecasts.

entrants will ameliorate physics market problems; according to the standard calculations, increases in supply will exacerbate the potential imbalance of the 1980's due to the likely fall in demand for faculty. While different changes in exogeneous variables such as a sudden sharp increase in federal *R&D* spending on energy would significantly invalidate our forecasts, it is evident from Table 7 that economic responsiveness has a first-order effect on the market and should be taken into account in manpower analysis.

III. Summary

The principal findings of this study can be summarized briefly:

1) The number of physics students has varied greatly in the postwar period, declining in the recent collapse of the science market after a decade of rapid increase. These changes result from economic responses to salary or job opportunity incentives, with fairly large elasticities of supply for physics salaries ranging from about unity in the short run to 3 or 4 in the long run at the graduate level, and from 2 in the short run to 4 in the long run for B.S. recipients.

2) Salaries of physicists have fluctuated significantly relative to other wages and salaries, rising in the post-Sputnik boom period and falling rapidly in the late 1960's and early 1970's. The bulk of the change in salaries can be attributed to *R&D* expenditures and the *R&D* policies of the federal government, with changes in supply having a smaller, though significant, effect.

3) Developments in the physics market, as in several other high level labor markets, appears to be well described by a simple recursive model which exploits the time delay in supply due to lengthy training programs. As a result of the recursive structure, large numbers of physics graduates create a market setting likely to reduce enrollments and future degrees in the field in accord with the cobweb scenario.

4) In the absence of a sharp increase in demand in the near future, market adjustments, especially on the supply side, are likely to lead to far fewer Ph.D.s in physics than indicated by standard forecasts. While the market still faces the major shock of a demographic-induced decline in demand for faculty in the 1980's, supply and salary adjustments will provide an important buffer to the decline.

REFERENCES

K. Arrow and W. Capron, "Dynamic Shortages and Price Rises: The Engineer-Scientist Case," *Quart. J. Econ.*, May 1959, *73*, 292–308.

G. Cain, R. Freeman, and L. Hansen, *Labor Market Analysis of Engineers and Technical Workers,* Baltimore 1973.

A. Cartter, "Scientific Manpower Trends for 1970–1985," *Science,* Apr. 1971, *172,* 132–40.

G. Chow and A. Lin, "Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Series," *Rev. Econ. Statist.,* Nov. 1971, *53,* 372–75.

J. Durbin, "Testing for Serial Correlation in Least-Squares Regression When Some of the Regressors are Lagged Dependent Variables," *Econometrica,* May 1970, *38,* 410–21.

R. Freeman, *Labor Market for College-Trained Manpower,* Cambridge 1971.

———, (1972a) "Labor Market Adjustments in Psychology," *Amer. Psychologist,* May 1972, *27,* 384–93.

———, (1972b) "The Science Manpower Market in the 1970's," unpublished paper, Harvard Univ., 1972.

———, "Autoregressive Patterns in Degrees: Evidence of Endogenous Adjustment Cycles?," unpublished paper, M.I.T. Center for Policy Alternatives, 1973.

———, "Legal Cobwebs: The Changing Market for Lawyers," *Rev. Econ. Statist.,* forthcoming.

——— and D. Breneman, "Forecasting the Ph.D. Labor Market: Pitfalls for Policy," Nat. Board Grad. Educ., Washington 1974.

M. Friedman, "The Interpolation of Time Series by Related Series," *J. Amer. Statist. Assn.,* Dec. 1962, *57,* 729–57.

Z. Griliches, "Specification Bias in Estimates of Production Functions," *J. Farm Econ.,* Feb. 1957, *36,* 8–20.

J. Muth, "Rational Expectations and the Theory of Price Movements," *Econometrica,* July 1961, *29,* 315–35.

M. Nerlove, "Adaptive Expectations and Cobweb Phenomena," *Quart. J. Econ.,* May 1958, *72,* 227–40.

D. Wolfe and C. Kidd, "The Future Market for Ph.D.'s," *Science,* Aug. 1971, *173,* 784–93.

American Institute of Physics, (1968–72a) *AIP Graduate Student Surveys,* nos. R-207.1–207.5, New York 1968–72.

———, (1968–72b) *Survey of Physics Bachelor's Degrees Recipients,* no. R211.1–211.4, New York 1968–72.

———, *Physics and Astronomy Enrollments and Degrees in U.S.,* no. R-151.10, New York, Mar. 1973.

College Placement Council, *(CPC)* "Men's Salary Survey," Bethlehem, Pa., 1960–1973.

Federal Interagency Committee on Education, Student Support Group, *Report on Federal Predoctoral Student Support,* Washington 1970, rev. 1972.

National Science Foundation, (1972a) "First-Year, Full-Time Graduate Science Enrollment Continues to Decline," Washington 1972.

———, (1972b) *National Pattern of R&D Resources,* 1953–1972, Washington 1972.

———, Reports of the National Register of Scientific and Technical Manpower, American Science Manpower, Washington 1954–55, 1956–58, 1960, 1962, 1964, 1966, 1968, 1970.

U.S. Atomic Energy Commission, *Annual Financial Federal Funds for Research Development,* Vol. 21, var. eds., Washington 1959–72.

U.S. Bureau of the Census, *Current Population Reports,* Series P-60, var. issues, Washington 1948–72.

———, *Statistical Abstract of the United States,* var. eds., Washington 1948–73.

U.S. Department of Commerce, *The National Income and Product Accounts of the U.S., 1929–1965,* Washington 1966.

———, *Survey of Current Business:* July eds., Washington 1966–72.

U.S. Department of Labor, *Employment, Education, and Earnings of American Men of Science,* Bull. 1027, Washington 1951.

U.S. Federal Security Agency, *Manpower Resources in Physics 1951,* Washington 1952.

The Balance of Payments and Money Supply Under the Gold Standard Regime: U.S. 1879-1914

By BIJAN B. AGHEVLI*

This paper analyzes the interaction between the real and monetary sectors via the balance-of-payments adjustment mechanism. An explicit theoretical model is developed and its coefficients are estimated (using two-stage least squares) for the United States during the period 1879–1914. The model is then synthesized to evaluate the views of Phillip Cagan, Milton Friedman and Anna Schwartz on one hand, and Jeffrey Williamson, Robert Mundell, and Moses Abramovitz on the other.

The period 1879–1914 in American economic history is of particular interest as it provides an excellent testing ground for theories pertaining to balance of payments and money supply. In this era the *U.S.* economy developed rapidly under a gold standard regime with little direct government intervention. There are conflicting interpretations of the events of this period with regard to the interaction between the monetary and real sectors.

Cagan and Friedman and Schwartz regard the variations in the money supply as independent of the changes in the real sector. Their view can be best summed up in the following passage from Friedman and Schwartz:

> One thing is abundantly clear from our narrative. Monetary changes have in fact often been independent, in the sense that they have often not been an immediate or necessary consequence of contemporaneous changes in business conditions.
>
> The clearest example is perhaps the monetary expansion from 1897 to 1914, which was worldwide and reflected an increased output of gold. The increased output of gold was partly a consequence of earlier decades of declining prices, which encouraged gold production, and so speaks also for a mutual interaction between monetary and economic changes. But clearly the monetary expansion cannot be attributed to the contemporary rise in money income and prices. By itself, the rise in money income and prices made for reduced output of gold in the world at large and for an outflow of gold from any single country in a gold-standard world. If the common movement of money and income was not purely coincidental, the direction of influence must run from money to income. [p. 686]

Cagan and Friedman and Schwartz, however, ignore the impact of changes in business conditions on the money supply via the balance of payments. In their analysis, they emphasize only the links between the balance on current account and prices, and treat the flows of capital along with the flow of specie as the passive part of the balance of payments. Thus, they ignore the effects of domestic real factors on the balance on current and capital account.

On the other hand, Williamson (1964) argued that variations in the money supply were linked to variations in the real sector through the balance of payments. He proposed that the variations in the balance of payments of the United States for the period 1879–1900 were dominated by capital movements, which in turn were caused by different rates of growth in the United States and Great Britain. Thus, the Williamson thesis specifies causation to run from rapid pace of development in the United States and more profitable investment opportunities to placement of securities in London, a capital surplus and gold inflow. Mundell's view of the relationship among the real sector, balance of payments and money supply is also along the lines proposed by Williamson. Mundell states:

---

* International Monetary Fund. I am indebted to George Borts, James Hanson, Jerome Stein, and Moses Abramovitz, who contributed many valuable comments and suggestions.

> The procyclical variation of interest rates in the United States causes after allowance for cycles abroad, a procyclical movement of the capital account which is usually associated with an anticyclical fluctuation of the trade-balance surplus. This appears to have been the historical pattern for the United States whenever it was not offset by cycles in the rest of the world. The capital account has dominated the trade balance procyclically and led to American payments surpluses in booms and deficits in depressions (the post-devaluation years, 1936–40, are an exception). Because the demand for money is procyclically strong it is associated with an excess supply of securities or goods, leading to a surplus on capital account, the trade balance, or both. [p. 7]

The relation between income, money supply, and balance of payments is explored in more detail by Abramovitz. He maintains that: "A more satisfactory treatment must recognize not only the impact of income change on the current and capital accounts; but it must also incorporate the effects of exogenous changes in the current and capital accounts on income and so on one another" (1973, p. 6). Abramovitz proposes that under a gold standard, the growth of money income and money supply is only possible through an *ex ante* (but not *ex post*) tendency in the economy to yield a balance-of-payments surplus. He is also in close agreement with the view that the developments in the real sector and domestic prices of real and financial assets had a strong impact on the balance of payments and consequently on the money supply. The main issues to be resolved, then, are the extent to which variations in the real sector influenced the trade and capital accounts, and whether the procyclical variation of capital inflows dominated the anticyclical variations of the trade balance.

In the following sections, an econometric model is constructed in which the relationships for the real and monetary sectors are specified. By estimating the coefficients of the model it is possible to determine to what extent the changes in the monetary sector were caused by the changes in the real sector. While the results support the view that capital flows move procyclically, they contradict the proposition that the procyclical movements of capital inflows dominated the anticyclical movements of the trade balance for the period under study.

I. The Model

In this section a two-country model is developed where each country produces one good. Goods can be exported or used for domestic consumption and business investment. Prices are determined by the conditions of equilibrium in the home and foreign goods markets. International movement of securities (bonds) is free, and bonds are assumed to be perfect substitutes regardless of nationality. As a result, there is a single rate of interest r which is determined by the condition of equilibrium in the world bond market.

The following notation is used for the two countries. The subscript f denotes the corresponding variable referring to the foreign country.

Y = real income
C = real consumption

I = real investment
S = real saving
P = price level
R = level of international reserves
r = nominal rate of interest
B_c = balance of trade
B_k = balance on capital account
M^D, (M^S) = nominal stock of money demanded (supplied)
XP = nominal exports

The following adjustment process, then, is assumed for the home and foreign price levels, world rate of interest, and level of international reserves. The operator D denotes time derivative and the coefficient λ_i is the speed of adjustment.

(1) $DP=\lambda_1$ {excess demand in the home country's goods market}
$$=\lambda_1\{IP-SP+XP-X_fP_f\}$$

(2) $DP_f=\lambda_1$ {excess demand in the foreign country's goods market}
$$=\{I_fP_f-S_fP_f+X_fP_f-XP\}$$

(3) $Dr=\lambda_3$ {excess supply in the world bond market}
$$=\{IP-SP+M^D-M^S+I_fP_f-S_fP_f+M_f^D-M_f^S\}$$

(4) $DR=$ {home country's balance-of-payments surplus}
$$=\{XP-X_fP_f+B_k\}$$

Derivation of the excess demand functions for home and foreign goods is quite straightforward.[1] Excess demand for bonds is derived by assuming that individuals save either by purchasing newly issued bonds or by adding to their money balances. Moreover, it is assumed that individuals adjust their holdings of money to desired levels by the purchase and sale of bonds, not goods. Excess supply of bonds in the world can be written then as the sum of excess of investment over savings and excess demand for money in each country.[2] (See equation (3).)

The level of international reserves will adjust as long as there is nonzero balance of payments according to equation (4). It should be noted that as long as any market is out of equilibrium, the rest of the system will also be thrown out of equilibrium. The balance of payments, therefore, will always reduce to zero at equilibrium. Furthermore, equilibrium in the goods and bond markets concurrent with zero balance of payments implies that excess demand for money in each country is also equal to zero. Consequently this system satisfies Walras' law (aggregate budget constraint), allowing the bond market to be dropped from view and replaced by the money market. Specification of the bond market, however, facilitates explanation of the way securities are traded between countries, and leads to a definition of the balance on capital account.

II. Balance on Capital Account

A country's balance on capital account is determined by its excess sale of securities to foreigners over its purchase of foreign securities. It is assumed that bonds are perfect substitutes for each other independent of the issuing country, and they are traded freely in what is called a world bond market.[3] In this framework, rates of interest at home and abroad will move together and interest rate differentials will not be observed. Thus, to the extent that the world interest rate is determined endogenously, the balance on capital account cannot be specified as a function of the interest rate differentials between the countries.

---

[1] For derivation of these excess demand functions and the stability analysis of the dynamic system, see the author and Borts.

[2] Assume that net flow demand for money, $m^d=K(M^D-M^S)$. For simplicity and with no loss of generality set the speed of adjustment K equal to unity. Excess demand for money is, then, given by the expression (M^D-M^S).

[3] It could have been assumed that the rate of return of foreign bonds for domestic residents is lower than r due to a "transfer" cost. However, as long as this cost is fixed and residents of each country view the other country's bonds to be as riskless as their own, this will not affect the analysis.

The domestic demand for foreign long-term capital inflows B_k^d is determined by the excess of new securities issued at home over the domestic financial demand. On the other hand, the foreign supply of capital inflow B_k^s is determined by the excess of foreign financial demand for securities over the foreign securities issued abroad. Thus, we can write:

$$(5)\quad \begin{aligned} B_k^d &= IP - SP + M^D - M^S \\ B_k^s &= -I_fP_f + S_fP_f - M_f^D + M_f^S \end{aligned}$$

Excess supply of bonds in the world bond market $\bar{E}$ can then be written as:

$$(6)\qquad \bar{E} = B_k^d - B_k^s$$

First, assume that the world bond market is in equilibrium ($\bar{E}=0$). This implies that the domestic demand for foreign capital inflows B_k^d matches the foreign supply of capital outflows B_k^s. That is, $\bar{E}=0$ implies $B_k^d=B_k^s$. In this case, the home country's balance on capital account B_k (B_k which is equal to the negative of the foreign country's balance on capital account) is given by either B_k^s or B_k^d.

$$(7)\quad \begin{aligned} B_k &= IP - SP + M^D - M^S \\ &= -I_fP_f + S_fP_f - M_f^D + M_f^S \end{aligned}$$

Depending on the lending or borrowing conditions of the home country at initial equilibrium, B_k could be negative, zero, or positive.

It is so far assumed that the world bond market is in equilibrium. The above formulation of the balance on capital, however, does not apply when the world bond market is in disequilibrium. In this case, there will be an excess supply (demand) for bonds in the world and $\bar{E}$ will be greater (less) than zero. Thus, the home country's demand for capital imports will exceed (fall short of) the foreign country's supply of capital. Clearly, as long as the world bond market is in disequilibrium, some individuals must remain unsatisfied. The question is, how do these unsatisfied demanders or suppliers of bonds distribute themselves between the two countries? Since it is assumed that people have free access to the bond market and they all regard all bonds as perfect substitutes, it is reasonable to conclude that unsatisfied individuals would be distributed at random between the two countries. Therefore, the home country's balance on capital account will be determined as a weighted average of the home country's demand for capital inflows and foreigners' supply of capital outflows. The weights would be determined by the size of the two economies and it will be assumed that once they are established they will remain unchanged.

Constraining the weights to add up to unity and denoting them by u and $(1-u)$, the home country's balance on capital can be written as:

$$(8)\quad \begin{aligned} B_k &= uB_k^d + (1-u)B_k^s \\ &= u[IP - SP + M^D - M^S] \\ &\quad + (1-u)[S_fP_f - I_fP_f + M_f^S - M_f^D] \end{aligned}$$

The above expression for the balance on capital account reduces to the previous formulation (equation (7)) at equilibrium. It should be noted that B_k could be a positive or negative function of the rate of interest depending upon the relative size of the two economies, as the following expression indicates.

$$(9)\qquad \frac{dB_k}{dr} = u\frac{dB_k^d}{dr} + (1-u)\frac{dB_k^s}{dr} \gtrless 0$$

In the above framework, the balance on capital account is ultimately determined by the equilibrium condition of the bond market. What, then, determines the long-term changes in capital flows? Once the system is in equilibrium the balance on capital account will change only as a result of shifts in investment, savings, or demand for money at home and abroad. The rate of interest always adjusts to eliminate any excess demand in the bond market. In this context, the relation between the rate of interest and capital flows is based on the nature of these exogenous shifts. Therefore, capital flows cannot be estimated as a function of the interest rate due to obvious identification problems. This could explain the failure of many empirical studies to identify a positive relationship between the interest rate and capital flows.

To verify the above analysis empirically, it is assumed that Great Britain was the only source of capital imports of the United States. This is not an unrealistic assumption as Great Britain was the main creditor of the world, and London was the major capital market in this period. Therefore, the British excess demand for bonds could be taken as a good proxy for the foreigners' supply of capital to the United States.

For the moment, the assumption is made that nominal investment $\bar{I}P$ is composed of two parts. The first part is denoted by $I(r)$ and represents that portion of the investment which responds to variations in the nominal interest rate. The second part, IP, represents that portion of investment which shifts exogenously due to growth factors. Also assuming that the money market is in equilibrium, domestic financial demand for securities is simply given by nominal savings. Then, writing the interest elastic portion of the *U.S.* investment $I(r)$ and the normal savings SP as linear functions of the interest rate and nominal income, and assuming that analogous relationships hold for the United Kingdom (and subscripted appropriately by f), the following relationships are derived:

$$(10) \quad I(r) = i_1 - i_2 r, \qquad i_2 > 0$$

$$(11) \quad SP = s_1 + s_2 r + s_3 YP \qquad s_2, s_3 > 0$$

Then

$$\begin{aligned}(12) \quad B_k^s &= \bar{I}P - SP = IP + I(r) - SP \\ &= (i_1 - s_1) - (i_2 + s_2)r + IP \\ &\quad - s_8 YP\end{aligned}$$

and analogously

$$\begin{aligned}(13) \quad B_k^d &= -\,(\bar{I}_f P_f - S_f P_f) \\ &= -\,[I_f P_f + I_f(r) - S_f P_f] \\ &= -\,(i_1' - s_1') + (i_2' + s_2')r \\ &\quad - I_f P_f + s_3' Y_f P_f\end{aligned}$$

Assuming that the rate of interest varies to equilibrate the bond market, the following reduced form equation for capital flows is derived by eliminating the interest rate.

$$\begin{aligned}(14) \quad B_k &= b_1 + b_2 IP - b_3 YP - b_4 I_f P_f \\ &\quad + b_5 Y_f P_f\end{aligned}$$

where $0 < b_2, b_3, b_4, b_5 < 1$.

In order to estimate the above equation,[4] Abramovitz's index of nominal construction for the United States and Alexander Cairncross' British investment series are used as proxies for IP and $I_f P_f$, respectively. It should be recognized that the Abramovitz construction series is not a completely satisfactory proxy for aggregate *U.S.* investment. In the absence of any other meaningful investment series however, the choice of this proxy is a forced one. Moreover, to the extent that railroad investment was the major vehicle for capital imports in this period and the movements of the Abramovitz index closely paralleled those of railroad investment, there is more justification for using the above series in estimating capital flows. As the following regression indicates, all of the parameters have the predicted signs and are significant at 1 percent confidence level, except the Y_fP_f term which has the predicted sign but is not significant. The t-values are given directly below the coefficients.

---

[4] Since separate data for long- and short-term capital are not available prior to 1900, B_k includes both categories for this period. For the description of other series used, see the Appendix.

$$(15)\quad \begin{aligned} B_k = {} & 346.1 + \underset{(3.1)}{\underset{(3.1)}{}} \underset{(3.1)}{0.088IP} - \underset{(2.9)}{0.024YP} \\ & + \underset{(0.25)}{0.042Y_fP_f} - \underset{(4.6)}{0.0053I_fP_f} \end{aligned}$$

$$\bar{R}^2 = .76 \qquad D.W. = 2.03$$

It should also be noted that the $U.S.$ and British investment series were not detrended or smoothed out by averaging procedures in order to guarantee that they reflected all the short-term oscillations in the pace of economic activity.[5]

III. The Balance on Trade Account

The balance on trade account is specified by introducing the export and import functions. Given that the United States' demand for imports was only a small fraction of the total world demand, it did not affect import prices appreciably. Thus, it is assumed that the $U.S.$ imports were determined according to the demand schedule and import prices were determined exogenously. Assuming that import goods are used for consumption and investment and specifying the demand for consumption goods as a linear function of nominal income, the demand for nominal imports ZP_z is estimated as follows:[6]

$$(16)\quad ZP_z = \underset{(6.4)}{2.49} + \underset{(3.6)}{0.097IP} + \underset{(2.2)}{.015YP}$$

$$\bar{R}^2 = .93 \qquad D.W. = 1.38$$

In the above formulation, it is assumed implicitly that investment influenced imports independently of income. It is argued that even though investment and income moved parallel to each other in the long run, their movements did not closely coincide in the short run. In fact, during business expansions, investment absorbs such a large share of output that imports rise by much larger amounts than the rise in output would suggest.

Next, the demand for and supply of exports are considered. Export prices, unlike import prices, are determined endogenously in order to equilibrate the foreign demand for and the domestic supply of exports. Supply of real exports X^S is specified as a positive function of the ratio of export prices P_x to the domestic price level P. Moreover, the export sector grew along with the rest of the economy due to the growth of capital stock, the labor force, and technological advancements. Using the level of the real output Y as a scale factor,[7] the supply of real exports is:

---

[5] Alternatively, Williamson (1963) has used Ulmer's series of net capital expenditures on American railroads $I'_{U.S.}$ as a "pull factor" and the Cairncross' series of British home investment as a "push factor" to explain the inflows of capital K. Using trendless annual data, smoothed by a moving five-year average, he obtained the following regression results for the period 1871–1914 (standard errors are given below the coefficients):

$$K = 914 + \underset{(0.20)}{0.03\, I'_{U.S.}} - \underset{(0.16)}{8.84\, I'_{G.B.}} \qquad R^2 = .65$$

Williamson's statistical methods of using five-year moving averages of detrended data were challenged by J. Ernest Tanner and Vittorio Bonomo. They maintained that "A simple five-year moving average may leave many cyclical influences in damped form and induce spurious correlation of its own" (p. 47). Tanner and Bonomo proceeded to remove the cyclical influence by averaging the data over the National Bureau reference cycles. They concluded that their evidence " . . . shows no positive correlation between the rate of economic growth and the rate of net capital inflow" (p. 51). To the extent that capital flows react to short-term oscillations in the investment series, however, we should not use any smoothing or averaging procedures which eliminate these oscillations.

[6] The Durbin-Watson ($D.W.$) statistic falls above the upper bound margin (i.e., $d_u = 1.37$) which implies there is no serial correlation in equation (18).

[7] Alternatively, the stock of capital, the labor force, or a time trend could have been used instead of the

$$(17) \qquad X^S = f_0(P_x/P)^{f_1} Y^{f_2} \qquad f_0, f_1, f_2 > 0$$

The demand for real exports depends on the level of foreign income Q and the terms of trade (i.e., the ratio of foreign prices P_z to export prices P_x). Denoting the elasticity of foreign demand for exports with respect to foreign income and the terms of trade by g_1 and g_2, respectively, the demand for real exports is:

$$(18) \qquad X^D = g_0 Q^{g_1}(P_z/P_x)^{g_2} \qquad g_0, g_1, g_2 > 0$$

Assuming that the elasticity of demand for real exports with respect to the terms of trade, g_2, is equal to unity, nominal demand for exports can be estimated as a function of foreign variables alone (equation (18)). That is, the nominal demand for exports becomes basically a reduced form equation and can be estimated independent of export prices in the simultaneous system. Constraining the coefficient of foreign prices to unity ($g_2 = 1$) gives the following results:[8]

$$(19) \quad Log\,(XP_x) = \underset{(10)}{-\,2.64} + \underset{(21)}{1.13}\,Log\,Q + Log\,P_z$$

$$\bar{R}^2 = .93 \qquad D.W. = 1.20$$

The above assumption of unit elasticity of export demand with respect to the terms of trade is a strong assumption and should be supported with some empirical evidence. In order to provide this, the foreign demand for real exports (equation (18)) is estimated simultaneously with the supply of real exports (equation (17)), using two-stage least squares method.[9] The following results are obtained for estimation of equation (18).

$$(20) \qquad Log\,X = \underset{(9.4)}{-\,2.63} + \underset{(19)}{1.13}Q + \underset{(2.5)}{0.96}\,Log\,(P_z/P_x)$$

As the above results indicate, the estimated value for coefficient g_2 is not significantly different from one. Thus, the empirical results provide us with strong justification for assuming g_2 is equal to unity.

IV. Demand for Money and Its Composition

This period of American economic history was characterized by a rapid expansion of the financial sector. The development of banking facilities had two important effects on the amount of private monetized transactions. The first effect was the educational element which helped to familiarize the public with the advantages of engaging in monetized transactions. As banking facilities spread to the rural areas, they helped to popularize money as a more convenient means of transactions as opposed to barter trade. In addition, rapid expansion of new banks decreased the opportunity cost of monetized transactions relative to barter trade, particularly as banks became more conveniently located for people to use their services.[10] The number of banks is used here as an index of banking development, and is denoted by N. The amount of real per capita monetized transactions T can then be written as a function of real per capita income (Y/L) and the number of banks N, according to the following equation.

---

income variable. The value of real income, however, seems to be a more reasonable choice as it indicates that the export sector grew along with the rest of the economy.

[8] The $D.W.$ statistic falls in the indeterminant range. Correction for serial correlation, however, does not change the results appreciably as the following corrected regression equation indicates.

$$Log\,(XP_x) = 2.37 + \underset{(5.7)}{1.07}\,Log\,Q + \underset{(12)}{}Log\,P_z;$$

$$\bar{R}^2 = .94 \qquad D.W. = 2.09$$

[9] For a description of the data, see the Appendix.

[10] See Robert Barro, James Hanson, and Luis Rosas.

$$(21) \qquad T = T_0(Y/L)^{q_2}N^{q_3}$$

$$\frac{dT}{d(Y/L)} > 0; \quad \frac{d^2T}{d(Y/L)^2} > 0 \text{ implies } q_2 > 1$$

$$\frac{dT}{DN} > 0; \quad \frac{d^2T}{dN^2} < 0 \text{ implies } 1 > q_3 > 0$$

The first derivative of T with respect to income and the number of banks should clearly be positive. The second derivative of T with respect to income is assumed to be positive to indicate that as people's income rises, they engage in proportionally more monetized transactions. On the other hand, the second derivative of T with respect to the number of banks is assumed to be negative to indicate the diminishing returns associated with this variable.

The next step is to assume that individuals hold certain amounts of real balances m which they replenish periodically by shifting their assets from interest-bearing assets (i.e., bonds) to money. There are two types of costs associated with such shifts. First there is an opportunity cost associated with the foregone interest payments on interest-bearing assets. Second, there are certain transfer costs involved in shifting from bonds to money. These transfer costs are of two kinds: a fixed cost which can be viewed as the foregone opportunity cost associated with the inconvenience of going to the broker to make the transfer and an explicit cost (i.e., "broker" fees) which is determined by the amount transferred. By assuming that each individual chooses his average real balances m in such a way to minimize the total cost, it can be shown[11] that m is given by the following relationship:

$$(22) \qquad m = P_0 r^{-P_1}(Y/L)^{P_2}N^{P_3}$$

where $\frac{1}{2} < P_1 < 1$; $\frac{1}{2} < P_2 \simeq 1$; $\frac{1}{2} < P_3 < 1$. The coefficients $-P_1$, P_2, and P_3 denote the elasticity of real balances per capita with respect to the rate of interest, real income per capita, and the number of banks, respectively. The following regression[12] indicates that the estimated coefficients conform significantly to the values hypothesized in equation (23).[13]

$$(23) \quad Log\left(m = \frac{M}{PL}\right) = \underset{(5.3)}{-\,2.87} - \underset{(8.8)}{0.83}\, Log\, r + \underset{(6.8)}{0.94}\, Log\,(Y/L) + \underset{(6.6)}{0.40}\, Log\, N$$

$$\bar{R}^2 = 0.99 \qquad D.W. = 1.46$$

It should be noted that the above estimated demand for money function is quite interest responsive (i.e., the estimated interest elasticity is 0.83). Thus, these results directly contradict Friedman and other monetarists who contend that demand for money is not very responsive to interest rates.

So far, individuals' demand for money has been derived irrespective of the form in which it is held (i.e., deposits or currency). The individual's desired deposit-currency ratio is determined by the relative attractiveness of deposits and currency. Holding balances in the form of deposits not only provides people with a positive rate of return (banks in this period paid interest on demand as well as time deposits), but is also a safer method of holding balances and provides a permanent record of transactions. Individuals are forced to hold certain amounts of currency in hand since checks are not always acceptable in transactions. As the real income per capita rises, however, the percentage of transactions which can be financed by checks would rise. Thus, the deposit-currency ratio would be positively correlated with per capita income.

---

[11] For a theoretical derivation of equation (25), see the author.

[12] For the description of data used, see the Appendix.

[13] The *D.W.* statistic falls above the upper bound margin (i.e., $d_U = 1.44$) which implies there is no serial correlation in equation (25).

The other important factor contributing to the rise in the deposit-currency ratio in this period was the development of the banking sector. The rise in the number of banks not only increased the demand for money by familiarizing people with the advantages of checking, but also induced a shift in balances from currency to deposits by making banking services more accessible. The deposit-currency ratio, then, is estimated as a log-linear function of real incomes per capita and the number of banks, where deposits and currency are denoted by D and C, respectively.[14]

$$(24)\quad Log\,(D/C) = \underset{(12)}{-\,8.3} + \underset{(3.2)}{0.88}\,Log\,(Y/L) + \underset{(4.0)}{0.53}\,Log\,N$$

$$\bar{R}^2 = 0.91 \qquad D.W. = 0.91$$

V. Supply of Money

In this section the supply of money is specified under a fractional reserve system. Using the identity derived by Friedman and Schwartz, supply of money can be formulated in terms of its three determinants; namely, high-powered money (H), deposit-currency ratio (D/C), and the deposit-reserve ratio (D/R).

$$(25)\quad M^S = H\left[\frac{D/R(1 + D/C)}{(D/R + D/C)}\right] \equiv KH$$

$$(26)\quad K = \frac{D/R(1 + D/C)}{D/R + D/C}$$

TABLE 1

Proximate Determinants of the Money Supply	Changes in Money Supply Attributed to Changes in Indicated Determinants as a Fraction of Total Change (1879–97)	(1897–1914)
High-powered money	.66	.64
Deposit-currency ratio	.24	.13
Deposit-reserve ratio	.06	.17
Interaction of ratios	.04	.05

Friedman and Schwartz were able to identify the extent that each of the above determinants of the money supply contributed to its variations. Their results for the periods 1879–97 and 1897–1914 are summarized in Table 1.

The most important source of change in the money supply was high-powered money. Changes in high-powered money were, in turn, caused primarily by changes in the monetary gold stock (approximately 68 percent). Changes in the monetary gold stock were due to (a) the flow of international gold, and (b) the excess of domestic gold production over consumption (i.e., jewelry, arts and crafts, hoarding, etc.). Thus, high-powered money H can be written according to the following identity.

$$(27)\quad H = H_{t-1} + E + (G_p - G_c) + B$$

where:

H_{t-1} = high-powered money in the previous period
E = changes in nongold components of high-powered money (i.e., silver, national bank notes, treasury notes, etc.)
G_P = domestic gold production
G_c = domestic gold consumption
B = balance of international payments

Assuming that domestic gold consumption is determined exogenously, and letting

$$(28)\quad H_0 \equiv H_{t-1} + E - G_c$$

---

[14] The $D.W.$ statistic indicates presence of serial correlation. Correction of serial correlation, however, does not change the results significantly, as the following corrected regression indicates:

$$Log\,(D/C) = \underset{(6.4)}{-8.3} + \underset{(1.8)}{0.83}\,Log\,(Y/L) + \underset{(2.6)}{0.53}\,Log\,N;$$

$$\bar{R}^2 = 0.91 \quad D.W. = 2.07$$

We can write:

$$H = H_0 + G_P + B \tag{29}$$

The domestic production of gold G_p is influenced by technological advancements and new discoveries in the gold mining sector. Following the discovery of rich mines in California in 1848, there were no important discoveries of gold in the United States until 1887. Thus, the *U.S.* gold production declined slowly all through the 1880's. Around 1888, however, the production of gold started to rise again due to innovations in the purification of gold ore (i.e., use of chlorination process and cyanide of potash in the extraction process).[15] Moreover, one would expect to see a negative relationship between the domestic price level and gold production. As the price level rises (falls), the supply of gold should fall (rise) as its real price in terms of other goods falls (rises), (i.e., the nominal price of gold is fixed by the government). In order to substantiate the above relationship, the period is split into two intervals and a trend term is used to capture the secular fall in production due to the depletion of resources prior to 1890, and the subsequent rise due to the new discoveries and technological advancements after 1890. Estimation of gold production as a function of price level and a trend gives the following results:[16]

$$(30)\quad Log\, G_P = -\underset{(3.6)}{2.06} - \underset{(2.7)}{2.96}\, Log\, P - \underset{(3.2)}{0.057t} \qquad (1879\text{–}89)$$

$$\bar{R}^2 = 0.52 \qquad D.W. = 1.13$$

$$Log\, G_P = \underset{(2.3)}{0.98} - \underset{(4.8)}{2.14}\, Log\, P + \underset{(12.2)}{0.077t} \qquad (1890\text{–}1914)$$

$$\bar{R}^2 = 0.93 \qquad D.W. = 0.52$$

In addition to the variations in the gold stock, variations in the deposit-currency and deposit-reserve ratio also contributed to the rise in the money supply. Inasmuch as the changes in the deposit-reserve ratio were marginal, no attempt is made to explain them and it is assumed that this ratio is determined by the government and the banking sector exogenously. The deposit-currency ratio, however, is determined endogenously according to equation (24) derived in the last section.

VI. The Simultaneous Model

In the previous sections, all the structural equations have been estimated using single equation techniques. In order to guarantee that these equations are not misspecified and the simultaneous bias does not invalidate the results, the complete system is reestimated using two-stage least squares method. The system is composed of six structural equations which in addition to two identities determine all eight endogenous variables of the system, namely, XP_x, ZP_z, B_k, M^D, M^S, D/C, G_p, and P. The estimation results are given below:

$$(31)\quad Log\,(XP_x/P_x) = \underset{(10)}{2.64} + \underset{(21)}{1.13}\, Log\, Q$$

$$(32)\quad ZP_z = \underset{(6.3)}{2.49} + \underset{(3.6)}{0.099}IP + \underset{(2.2)}{0.015}YP$$

$$(33)\quad B_k = \underset{(3.1)}{343} + \underset{(3.0)}{0.087}IP - \underset{(2.9)}{0.024}YP - \underset{(4.6)}{5.29}I_fP_f + \underset{(0.28)}{0.047}Y_fP_f$$

$$(34)\quad Log\frac{M}{PL} = \underset{(5.3)}{2.87} + \underset{(6.8)}{0.94}\, Log\frac{Y}{L} + \underset{(6.6)}{0.40}\, Log\, N - \underset{(8.8)}{0.84}\, Log\, r$$

---

[15] See Cagan, p. 59.

[16] The *D.W.* statistic indicates presence of serial correlation. Since the trend term explains most of the variation (see Table 1), equations (30) have not been corrected for serial correlation.

$$\text{(35)}\quad Log\, D/C = \underset{(12)}{8.3} + \underset{(3.2)}{0.88}\, Log \frac{Y}{L} + \underset{(3.2)}{0.53}\, Log\, N$$

$$\text{(36a)}\quad Log\, G_p = -\underset{(3.6)}{2.06} - \underset{(2.7)}{2.96}\, Log\, P - \underset{(12)}{0.057t} \qquad (1878\text{–}1889)$$

$$\text{(36b)}\quad Log\, G_p = \underset{(2.4)}{1.01} - \underset{(4.9)}{2.12}\, Log\, P + \underset{(12.4)}{0.076t} \qquad (1890\text{–}1914)$$

$$\text{(37)}\quad M^S \equiv (H_0 + G_p + XP_x - ZP_z + B_k)K; \qquad K = \frac{D/R(1 + D/C)}{D/R + D/C}$$

$$\text{(38)}\quad M^D \equiv M^S$$

In the above simultaneous system, the capital flow equation (33) was derived by assuming that the domestic interest rate moved parallel to the foreign rate[17] and cleared the bond market. Prices, on the other hand, are determined by the equilibrium condition in the money market. That is, for given supply of nominal money and demand for real balances, prices adjust in order to equate the supply and demand for nominal money. For estimation purposes, it was simpler to specify the bond and money markets and eliminate goods market by invoking Walras' law. This arbitrary choice of markets, however, should not have any effects on the results.

Three of the structural equations in the above system are reduced form equations since they are exclusively functions of exogenous variables; namely, equations (31), (34), and (35). Moreover, the break in the gold production equation makes the number of observations too few for simultaneous estimation of equation (36a). Therefore, equations (31), (34), (35), and (36a) are reproduced here from earlier ordinary least squares results. The above system of equations is not a linear one. Nevertheless, the two-stage least squares method can be used to derive consistent estimators for a simultaneous system which is non-linear in the endogenous variables but linear in the parameters.[18] As it can be seen, the simultaneous estimation of equations (32), (33), and (36b) hardly changes the value of the coefficients which gives us more confidence in their validity.

It is useful to examine the importance of individual variables in terms of their contributions to explain the variations in the dependent variables. Clearly, the size of estimated coefficients of individual regressors are no measure of importance as they are dependent on the units of measurement used. Thus, the so-called "*beta* coefficients"[19] are used here, which are derived by multiplying the coefficients of individual regressors by the ratio of their standard deviation to the standard deviation of the dependent variable, in each equation. The *beta* coefficients are standardized so that the sum of their absolute values in each question adds up to 100. These coefficients indicate roughly the relative importance of different variables in terms of their contributions to the variations in each regression. The results are summarized in Table 2, where the coefficients are listed from left to right in order of their importance.

As can be seen in this table, the rise in exports is mostly due to the rise in the value of world imports; while the rise in the *U.S.* imports is explained by the rise in domestic investment and income. Oscil-

---

[17] For a detailed analysis of the nature of interdependence between the *U.S.* and British interest rates, see Oskar Morgenstern.

[18] See Harry Kelejian.

[19] See Arthur Goldberger, p. 197.

TABLE 2—*Beta* COEFFICIENTS OF INDIVIDUAL REGRESSORS

Exports (XP_x)	Q	(P_s/P_x)		
	96	4		
Imports (ZP_z)	IP	YP		
	51	49		
Balance on capital account (B_k)	IP	I_fP_f	YP	Y_fP_f
	64	−21	−14	1
Real balances per capita (M/PL)	N	(Y/L)	r	
	59	32	−9	
Deposit-currency ratio (D/C)	N	(Y/L)		
	72	28		
Domestic gold production (G_p)	T (trend)		P	
1879–89	98.6		−1.4	
1890–1914	63		−37	

lations in the balance on capital account are to a large extent due to the changes in the domestic and foreign investments and to a lesser extent due to the changes in domestic and foreign income. The rise in the number of banks is the most important factor contributing to the rise in demand for real balances and deposit-currency ratio. This illustrates the importance of the banking development and its effects on the monetary sector. The rise in income per capita is also an important factor resulting in the rise of the demand for money and deposit-currency ratio.

VII. Synthesis of the Model

The results of the model can now be compared with those of Cagan and Friedman and Schwartz on one hand, and Williamson, Mundell, and Abramovitz on the other. In order to evaluate the above explanations of the events of the period, determinants of changes in the money supply are examined separately. Variations in the money supply are induced primarily by changes in the deposit-currency ratio and the monetary gold stock. The latter is caused by domestic gold production and the flow of international gold. The deposit-currency ratio is estimated as a function of per capita income and the number of banks, which can be considered an index of banking development. Domestic gold production is estimated as a function of the domestic price level and a trend factor to represent discoveries and technological advancements in the gold producing sector. The flow of international gold is determined by the balance on current account and the flow of capital. The latter is estimated as a function of domestic and foreign investment and income. Furthermore, nominal imports are estimated as a function of nominal investment and income, whereas nominal exports are estimated as a function of the volume of world imports.

The causes of changes in the above determinants of the money supply can be categorized into what might be called short-term and long-term. These categories are defined and analyzed separately.

A. *Short-Term Variations of the Money Supply*

Short-term changes in the money supply were induced by cyclical changes in business activity. These variations in economic activity affected the monetary gold stock by inducing variations in domestic gold production, trade balance, and international flows of capital. The effects of business cycles on the domestic production of gold and the trade balance were induced through the goods market; whereas the impact on capital flows was induced through the bond market. During the upswing of the cycle, rapid growth of output raises investment and income simultaneously in such a manner that the gap between investment and savings increases procyclically. This occurs because the rise in savings due to the rise in income during the upcycles is not enough to finance the rise in investment. The resulting gap be-

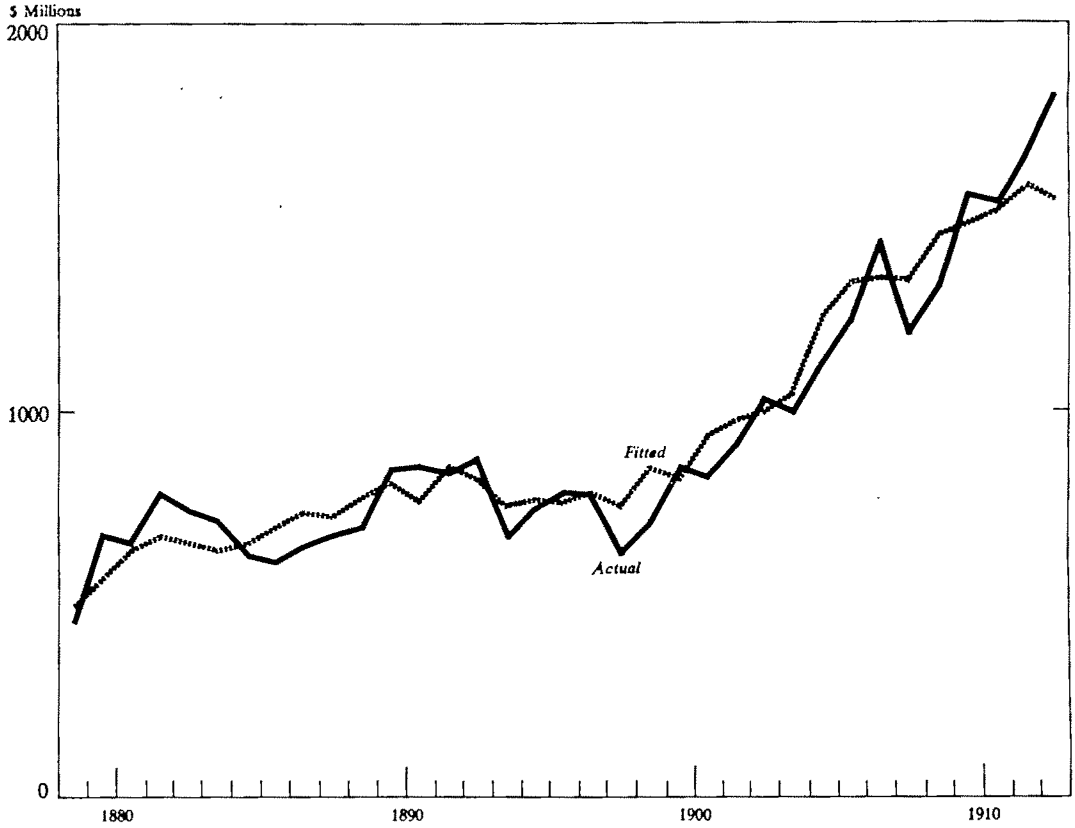


FIGURE 1. IMPORTS

tween domestic investment and savings creates an excess demand for goods and an excess supply of bonds.

The initial rise in income and investment raises imports by directly increasing the demand for consumer and investment goods (equation (32)). Moreover, during business upswings, the inflationary gap between investment and savings gives rise to an increase in domestic prices, which will ultimately eliminate the excess demand for goods. The resultant rise in prices will in turn change the terms of trade in favor of the home country, and therefore stimulate the demand for imports even further. Thus, assuming that export demand does not change (foreign demand for exports is determined exogenously), the trade balance moves anticyclically causing anticyclical variations in the monetary gold stock. Moreover, the rise in prices reduces the commodity value of gold and dampens its production (equations (36a) and (36b)), which in turn reinforces the anticyclical effects of the trade balance on the monetary gold stock.

Next, the effects of business cycles on the international capital flows are considered through the bond market. As already mentioned, the excess of domestic investment over savings during the upswings creates an excess supply of bonds forcing up the rate of interest. The rise in the interest rate eliminates the excess supply of bonds by (a) attracting capital inflows, (b) reducing the demand for money, and (c) decreasing the interest elastic portion of investment.

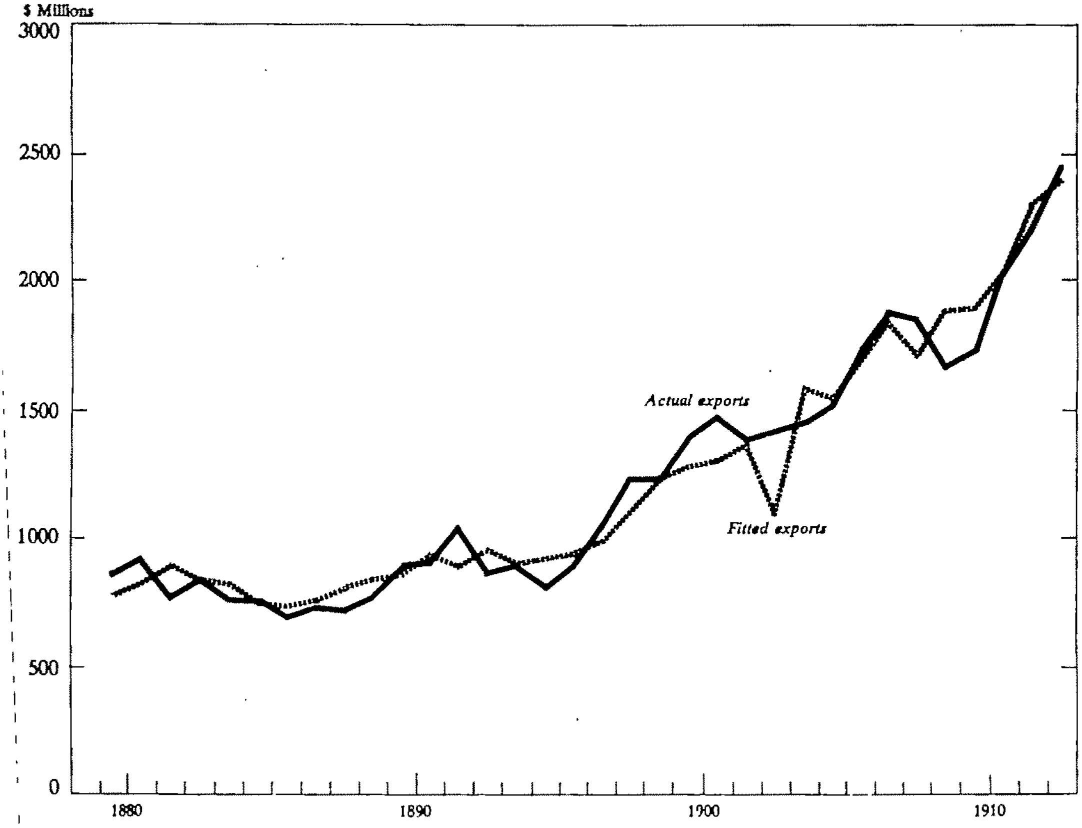


FIGURE 2. EXPORTS

The resultant effects of the shifts in income and investment on capital flows can be seen directly by considering the reduced form expression for capital inflows (equation (33)). According to this equation, an upward shift of one unit of investment, *ceteris paribus*, will induce a rise of 0.087 units in capital inflows. To the extent that income also rises during the upcycles, however, it helps to finance some of the rise in investment. Nevertheless, during the upswings the rise in savings is not enough to finance increased investments and the net effect is an inflow of capital and gold.

The above results confirm the first hypothesis set forth by Williamson and Mundell which states that allowing for cycles abroad, the capital inflows of the United States moved procyclically. The results do not confirm Williamson and Mundell's second hypothesis which claims that these procyclical movements of capital inflows dominated the anticyclical movements of imports. This second hypothesis can be easily tested by comparing the reduced form equations of capital flows (equation (33)) and demand for imports (equation (32)). Consider a simultaneous rise in investment and income such that the rise in investment exceeds the rise in savings and results in a capital inflow. First, consider the effects of the rise in investment by itself. Inasmuch as the estimated coefficient of investment in equation (32) exceeds the estimated coefficient of investment in equation (33), (.099 and .087, respectively), a rise in investment will increase capital inflows less than it

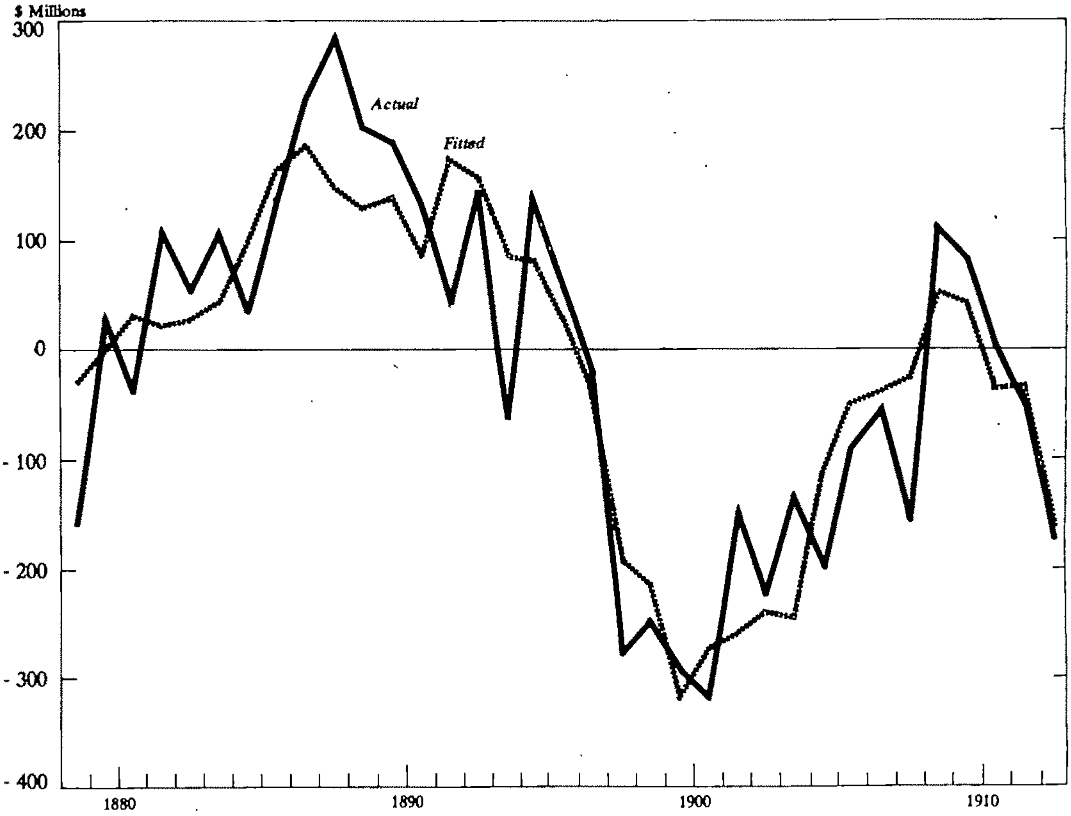


FIGURE 3. BALANCE ON CAPITAL ACCOUNT

raises imports. Thus, the resultant rise in capital inflows is more than compensated by the rise in imports, and the net effect of the rise in investment on the balance of payments is negative. Next, consider the effects of the rise in income which took place concomitantly with the rise in investment. The rise in income increases the demand for imported consumption goods and raises imports (see equation (32)). At the same time, the rise in income reduces the demand for foreign capital by raising domestic savings (see equation (33)). Consequently, the rise in income, *ceteris paribus*, will reduce the balance-of-payments surplus by reducing capital inflow and increasing imports. Therefore, the net effect of a simultaneous increase in income and investment on the balance of payments during an upcycle is negative. Thus, these empirical results indicate, contrary to Williamson and Mundell's second hypothesis, that the anticyclical movements of the trade balance dominated the procyclical movements of the balance on capital account and induced gold outflows.

So far the role of foreign factors on the balance of payments has been ignored. In fact, much of the *U.S.* capital flows in this period were caused by British business conditions. In particular, during 1897–1906, the United States was a large-scale net exporter of capital, a phenomenon which puzzled Friedman and Schwartz (see p. 142). During this period, net expenditure on British construction and railroads rose in excess of British savings. This factor, in addition to the existence of the excess of *U.S.* savings over investment, resulted in substantial capital outflows from the United States. In order to demonstrate how well the model explains the behavior of the balance of payments, the actual and fitted values of capital flows, imports, and exports are plotted in Figures 1, 2, and 3.

B. *Long-Term Variations of the Money Supply*

In the previous section it was established that the balance of payments responded to short-term variations in business activity and affected the money supply through the flow of international gold. It was argued, however, that the equilibrium of the dynamic system would imply a zero balance of payments in the long run. That is, prices and the rate of interest vary in such a way that the balance on capital account and balance of goods and services would counter each other. For empirical verification of the inverse relationship between the balance of goods and services *BGS* and the balance of capital accounts B_k, the following relationship is estimated.

$$B_k = \underset{(0.26)}{1.80} - \underset{(22)}{0.85}BGS \qquad (39)$$

$$\bar{R}^2 = .94 \qquad D.W. = 1.58$$

The above regression as well as Figure 4 illustrates that, in fact, the trade balance on goods and services and capital inflows moved opposite to each other in almost one-to-one fashion. Therefore, the net inflow of international gold was a rather small portion of the changes in the money

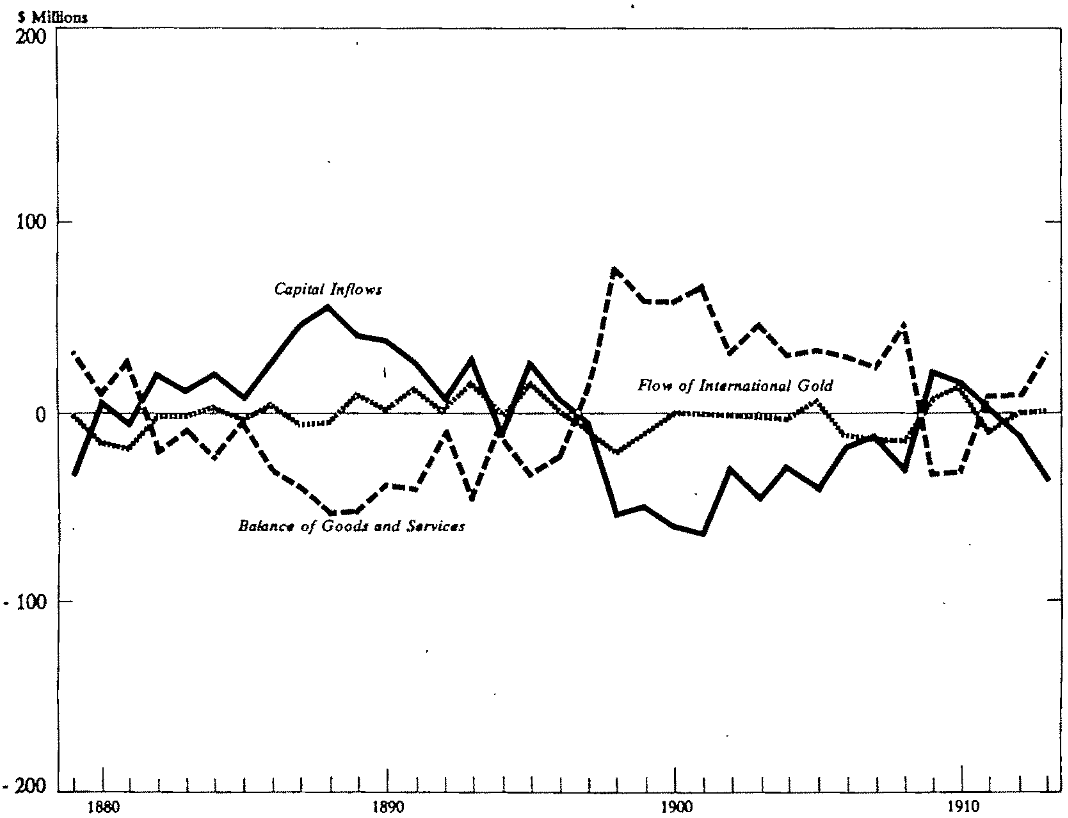


FIGURE 4

supply. Almost all of the additions to the monetary gold stock thus came from domestic production. Domestic gold production in turn was influenced by changes in domestic prices, new discoveries and technological advancements (equation (36)). The other source of long-term variations in the money supply was the deposit-currency ratio. This ratio rose over the period from a value of 2 to 7.5 due to the rise in per capita income and the number of banks (equation (35)).

We are now in a position to evaluate the Cagan and Friedman and Schwartz assumption of considering the money supply as totally exogenous. It has been shown that the rise of the money supply in the long run was caused by new discoveries of gold, improvement of gold extraction techniques, changes in domestic prices, rise of per capita income, and the increase in the number of banks. Of the above factors, the only one which is determined endogenously is the price level. Therefore, most of the long-term variations in the money supply were due to exogenous factors, a position which is consistent with Cagan and Friedman and Schwartz. As Abramovitz (1973) argues, however, it should be noted that while the rise in domestic gold production increased the money supply and satisfied the increased demand for money during this period, such increases in gold production were by no means a necessary element in the system to eliminate excess demand for money. In the absence of domestic gold production, there would have been an adjustment of *U.S.* prices; and hence of *U.S.* imports and exports, such that larger specie inflow would have compensated, at least in the short run, for smaller domestic gold production.

VIII. Conclusions

The different hypotheses set forth by Williamson, Mundell, Abramovitz, Cagan, and Friedman and Schwartz have been evaluated in the preceding sections. First, the thesis proposed by Williamson and Mundell that capital flows moved procyclically and dominated the anticyclical movements of the trade balance was tested directly. Going beyond the works of Williamson, a theory of capital flows was developed which is capable of explaining the oscillations of the balance on capital account surprisingly well. Although Williamson and Mundell were correct in proposing that capital flows moved procyclically, they were incorrect in arguing that these movements would dominate the trade balance movements.

In regard to the Cagan and Friedman and Schwartz position that the money supply was totally exogenous, the results are not as clear cut. It was established that in the short run, there were strong feedbacks from the real sector to the monetary sector through the change in prices and interest rate and their consequent effects on gold production and international gold flows. In the long run, however, the money supply was determined mostly by exogenous factors. Using a formal framework, these factors were identified as new discoveries and technological methods in gold production, rise in per capita income, and number of banks. Cagan and Friedman and Schwartz, however, give the rise in world gold production as the primary reason for the rise in the money supply. My results indicate that the rise in the money supply for this period was not entirely due to new discoveries and technological advancements in the gold producing sector. In fact, the rise in per capita income, financial developments, and changes in prices contributed significantly to the variations in money supply. Moreover, the model developed here is in agreement with Abramovitz's analysis which emphasizes the importance of balance-of-payments adjustment to clear the markets via variations in prices. It emerged that the money supply rose mostly due to the rise in domestic gold production. But this was an historical accident and in the absence of such rises in the gold production or other exogenous elements of the system, the balance of payments would react to changes in domestic prices and induce enough gold flows to eliminate excess demand for money.[20]

APPENDIX

Description of Data

(YP) = *U.S. Gross National Product in Current Prices* in millions of dollars, Simon Kuznets, p. 561, table 4–25, col. (3).

(Y) = *U.S. Gross National Product in 1929 Prices* in millions of dollars, Kuznets, p. 563, table 4–26, col. (3).

(P) = *Kuznets' Implicit Price Index of the United States.* Ratio of above two columns multiplied by 100.

(Y_fP_f) = *U.K. Net National Product in Current Prices* in millions of British pounds, Brian Mitchell, p. 367.

(I_fP_f) = *Cairncross Estimates of the U.K. Total New Construction* in millions of pounds, Cairncross, p. 169.

(QP_f) = *World Imports Excluding U.S. Imports* in millions of dollars, Ilse Mintz, p. 315, table c-1. The annual averages of the quarterly estimates given by Mintz are used.

(B_c+B_s) = *U.S. Balance of Goods and Services* in millions of dollars 1879–1900: Mathew Simon, p. 700, table 27, row 27. 1901–13: U.S. Bureau of the Census, p. 564, sum of col. U (182) and U (183).

(G_e) = *Exported Gold Stock of the United States* in millions of dollars 1879–1900: Simon, p. 700, table 27, row 28. 1901–13: U.S. Bureau of the Census, p. 537, col. U (6).

(B_k) = *U.S. Balance on Capital* in millions of dollars, 1879–1900: Simon, p. 700, table 27, row 30. 1901–13: Negative of the balance on goods and services plus the exported gold stock of the United States, given above.

(XP_x) = *Value of the U.S. Exports* in millions of dollars, Williamson (1964), p. 256, table B-1.

(ZP_z) = *Value of the U.S. Imports* in millions of dollars, Williamson (1964), p. 256, table B-1.

(P_x) = *U.S. Export Price Index,* 1913 = 100, Williamson (1964), p. 262, table B-4.

(P_z) = *U.S. Import Price Index,* 1913 = 100, Williamson (1964), p. 262, table B-4.

(r) = *Corporate Unadjusted Index Number of Yield of American Railroad Bonds,* U.S. Bureau of the Census, p. 656, col. X (332).

(N) = *Number of National Banks of the United States,* U.S. Bureau of the Census, p. 626, col. X (42).

(L) = *Estimated Population of the United States* in millions, U.S. Bureau of the Census, p. 7, col. A (2).

(G_p) = *U.S. Production of Gold* in millions of dollars, U.S. Bureau of the Census, p. 371, col. M (246).

(IP^1) = *Abramovitz Index of Construction in Current Prices* (1870–97) Abramovitz (1964), p. 142, series (3), Seg. I.

(IP^2) = *Abramovitz Index of Construction in Current Prices* (1889–1918) Abramovitz (1964), p. 142, series (3), Seg. II.

(IP) = *Total Construction in Current Prices:* To derive an estimate of nominal construction, the Abramovitz indexes of construction have been linked for the periods 1879–88 and 1889–1913. Then this index has been scaled (multiplied by 15.8) so it would roughly correspond to Kuznets' estimates of net capital formation. The Kuznets' estimates of net capital formation were not used as he removed all short-term variations by employing moving averages and other smoothing techniques.

---

[20] This view is also consistent with the monetary approach to the balance of payments developed by Harry Johnson.

(D) = *Time and Demand Deposits in Commercial Banks* in millions of dollars, Friedman and Schwartz, p. 740, table A-1, col. (4).

(C) = *Currency Held by the Public* in millions of dollars, Friedman and Schwartz, p. 740, table A-1, col. (1).

(M^S) = *Total Supply of Money* in millions of dollars, Friedman and Schwartz, p. 740, table A-1, col. (8).

(H) = *Total High-Powered Money* in millions of dollars, Friedman and Schwartz, p. 740, table A-1, col. (1).

REFERENCES

M. Abramovitz, "The Monetary Side of Long Swings in U.S. Economic Growth," memo. No. 14-6, Center Res. Econ. Growth, Stanford Univ. 1973.

———, *Evidence of Long Swings in Aggregate Construction Since the Civil War*, New York 1964.

B. B. Aghevli, "Effects of Banking Development on Demand for and Supply of Money," unpublished manuscript, Int. Monet. Fund 1973.

——— and G. H. Borts, "The Stability and Equilibrium of the Balance of Payments under the Fixed Exchange Rate," *J. Int. Econ.*, Feb. 1973, *3*, 1–20.

R. Barro, J. Hanson, and L. Rosas, "Financial Development, Transactions Costs, and Demand for Money," unpublished manuscript, Brown Univ. 1971.

G. H. Borts, "A Theory of Long-Run International Capital Movements," *J. Polit. Econ.*, Aug. 1964, *72*, 341–59.

P. Cagan, *Determinants and Effects of Changes in the Stock of Money, 1875–1960*, New York 1965.

A. K. Cairncross, *Home and Foreign Investment 1870–1913*, Cambridge 1953.

M. Friedman and A. J. Schwartz, *A Monetary History of the United States, 1867–1960*, Princeton 1963.

A. S. Goldberger, *Econometric Theory*, New York 1966.

H. Johnson, "The Monetary Approach to Balance of Payments Theory," in M. Conolly and A. Swoboda, eds., *International Trade and Money*, Toronto 1973.

H. H. Kelejian, "Two-Stage Least Squares and Econometric System Linear in Parameters but Nonlinear in the Endogenous Variables," *J. Amer. Statist. Assn.*, June 1971, *66*, 373–74.

S. Kuznets, *Capital in the American Economy: Its Formation and Financing*, Princeton 1961.

I. Mintz, *Cylical Fluctuations in the Exports of the United States Since 1879*, New York 1967.

B. R. Mitchell, *Abstract of British Historical Statistics*, Cambridge 1962.

O. Morgenstern, *International Financial Transactions and Business Cycles*, Princeton 1959.

R. Mundell, "The Dollar and the Policy Mix: 1971," Essays in Finance, no. 85, Princeton, May 1971.

M. Simon, "The United States Balance of Payments, 1861–1900," in *Trends in the American Economy in the Nineteenth Century*, Nat. Bur. Econ. Res. *Stud. in Income anl Wealth*, Vol. 24, Princeton 1960.

J. E. Tanner and B. Bonomo, "Gold, Capital Flows, and Long Swings in American Business Activity," *J. Polit. Econ.*, Jan./Feb. 1968, *76*, 44–52.

J. G. Williamson, *American Growth and the Balance of Payments, 1820–1913*, Chapel Hill 1964.

———, "Real Growth, Monetary Disturbances and the Transfer Process, 1879–1914," *Southern Econ. J.*, Jan. 1963, *29*, 167–80.

U.S. Bureau of the Census, *Historical Statistics of the United States, Colonial Times to 1957*, Washington 1960.

Investment Behavior, The Measurement of Depreciation, and Tax Policy

By Robert M. Coen*

Business income taxes are intended to tax income net of economic depreciation, that is, income net of the change in value of capital goods over the accounting period. If the effective rate of tax is to equal the statutory rate, then depreciation permitted for tax purposes must coincide with actual economic depreciation. When tax depreciation exceeds actual economic depreciation, the effective tax rate on true income is less than the statutory rate, and conversely.[1] To achieve tax equity and neutrality, the Treasury should strive to keep tax depreciation in line with economic depreciation; even if Treasury policy should seek to stimulate capital expenditures by reducing effective tax rates through liberalization of tax depreciation, a reasonable objective might be to reduce effective tax rates equally in all industries, which might require different degrees of liberalization in different industries. Efforts along these lines are greatly complicated, however, by the lack of adequate information regarding service lives and patterns of economic depreciation of capital goods in various industries.

The Treasury has relied primarily on engineering reports and special surveys of company experiences to acquire data on service lives and, to a more limited extent, on depreciation patterns; but the coverage has not been very complete, and the findings are necessarily somewhat suspect, since company representatives cannot be regarded as totally unbiased sources in such matters. Economists have attempted to measure economic depreciation principally by examining relationships between market prices of certain types of new and used assets—farm tractors, autos, and pick-up trucks, for example,[2] but the absence of active markets for many types of used industrial capital goods greatly limits opportunities for this type of analysis.

This study explores a new approach to the estimation of economic depreciation, an indirect method that attempts to infer patterns of economic depreciation from the behavior of actual capital expenditures in different industries. Section I presents a model of investment behavior that allows for alternative assumptions regarding service lives and capacity depreciation patterns. Capacity depreciation refers to losses in the productive capacity, or efficiency, of capital goods as they age. The model relates net investment to past changes in the desired, or optimal, stock of capital where the desired stock depends in a special way on the level of output, the price of output, and an implicit rental price of the capital input. Service lives and capacity depreciation patterns are shown to be central elements in the measurement of net investment and the implicit rental price of the capital input. To each service life and capacity depreciation assumption there corresponds a particular time-series for net investment and for the rental price. A range of alternative assumptions is employed in fitting the model to annual time-series data on equipment and structures expenditures in two-digit manufacturing industries. The statistical results are discussed in Section II. The service lives and capacity depreciation patterns that, among those tested, provide the best explanations of investment behavior are then translated into patterns of economic depreciation and compared with tax depreciation for each industry. This material appears in Section III. The concluding section summarizes the findings and indicates some of their important limitations.

---

* Associate professor of economics, Northwestern University. This research was supported by the Office of Industrial Economics, U.S. Treasury Department, under contract no. TIR-63-18. I want to thank Raymond Scheppach, Jr., Martin Marimont, and Jack Gottsegen for their aid in obtaining various data series. John Albrecht, Barry Cohen, Edward Hair, Seung Hoon Lee, Nancy Mueller, and Jon Sonstelie provided able assistance in various stages of the research. I am deeply indebted to Seymour Fiekowsky of the Treasury who gave me constant encouragement and much helpful advice from the formative stages to final completion of this study. I have also benefitted from the comments and suggestions of Robert Eisner. The analyses and conclusions contained in this study are my own and are not necessarily shared by any of these individuals or by the Treasury Department. A more detailed version of this paper is available on request from the author.

[1] A succinct proof of this proposition has been given by Paul Samuelson.

[2] See the studies of Zvi Griliches, Phillip Cagan, Frank Wykoff, Robert Hall, and Wolfhard Ramm. Paul Taubman and Robert Rasche have measured economic depreciation of office buildings by calculating changes in the discounted value of net rents as buildings age. Allan Mendelowitz has attempted to measure economic depreciation patterns in selected industries from estimated time profiles of earnings generated by capital expenditures in a given year.

I. Service Lives, Capacity Depreciation, and the Theory of Investment

A. *Gross Investment, Net Investment, and Replacement*

Firms acquire new capital goods both to expand productive capacity (net investment) and to replace losses in the productive capacity of existing capital goods (replacement investment). The partitioning of observed gross investment into net investment and replacement can only be approximated by making some assumptions about the service life and capacity depreciation pattern of capital goods. A widely employed assumption is that the productive capacity of capital goods declines geometrically (or exponentially) with age at a rate typically, though not necessarily, equal to twice the straight-line depreciation rate. This hypothesis implies that replacement is a constant fraction of the capital stock on hand at the beginning of each period. While this approximation is a convenient one in many respects, little empirical evidence has been advanced to support it.[3] Here I shall adopt a more general approach.

Let d_{ji} be the fraction of the original productive capacity of a capital good which is lost in the ith period after its acquisition according to the jth "mortality distribution." A mortality distribution is defined by a combination of an assumed service life and an assumed pattern of capacity depreciation.[4] Replacement in

---

[3] The studies of prices of new and used farm tractors, autos, and pick-up trucks referred to in fn. 2 do lend limited support to the hypothesis of geometric decline in productive capacity. Martin Feldstein and David Foot and Robert Eisner have undertaken more direct tests of the approximation, utilizing information on replacement expenditures from the McGraw-Hill surveys. Both studies lead to a rejection of the hypothesis that replacement is a constant fraction of capital stock. Dale Jorgenson (1971) has been critical of the Feldstein-Foot and Eisner studies, arguing that the capital stock estimates which they employ are inconsistent with a varying replacement rate.

[4] I assume that losses in productive capacity depend only on the time elapsed since the acquisition date. They are independent of the calendar date at which the capital good is acquired. It should be noted, however, that d_i may relate not only to deterioration in physical characteristics of the capital good but also to obsolescence associated with improvements in efficiency of new capital goods. For example, suppose that physical deterioration in the ith period following acquisition is h_i and that technical change embodied in new capital goods is occurring at the rate g per period. Then, following the aggregation theorem of Robert Solow for the case of a Cobb-Douglas production function with constant returns to scale, the capital stock for period t in efficiency units of period t capital goods can be defined as a weighted sum of current and past capital expenditures, namely,

$$K_t = (1 - h_0)I_t + \frac{1 - h_0 - h_1}{1 + g} I_{t-1} + \frac{1 - h_0 - h_1 - h_2}{(1 + g)^2} I_{t-2} + \dots$$

The weights are the same for all t, their numerators period t corresponding to the jth mortality distribution, assuming that $d_{j0}=0$, is given by

$$R_{jt} = \sum_{i=1}^{\infty} d_{ji} I_{t-i} \tag{1}$$

where I_{t-i} is real gross investment in year $t-i$, and net investment is given by

$$N_{jt} = I_t - R_{jt} \tag{2}$$

Given a sufficiently long historical series on gross investment, net investment series can be estimated for a wide range of mortality distributions. In this study, time-series on net investment in equipment in manufacturing industries have been constructed for 40 mortality distributions—8 service life assumptions generally ranging from 8 to 22 years, in increments of 2 years, and 5 capacity depreciation patterns. Net investment series for structures have been constructed for 35 mortality distributions—7 service life assumptions ranging from 20 to 50 years, in increments of 5 years, and 5 capacity depreciation patterns.[6] The assumed depreciation patterns, which are the same for both equipment and structures, are as follows:

1) A geometric decay pattern with the rate of decay set equal to $2/n$, where n is the assumed service life. For this pattern, $d_i=\delta(1-\delta)^{i-1}$, where $\delta=2/n$, and assets continue to contribute to productive capacity beyond the assumed service life. The meaning of "service life" is somewhat ambiguous in this case; strictly speaking, the choice of n should be viewed not as specifying different service lives but different rates of geometric decay. Because capacity depreciation extends over an infinite horizon according to this pattern, I shall refer to it in brief as *GD-INF*.

2) A geometric decay pattern with the rate of decay equal to $2/n$, but truncated so that assets do not contribute to productive capacity beyond the end of the service life. This pattern shall be referred to in brief as *GD-FIN*.

3) A "sum-of-the-years-digits pattern" for which $d_i=(n+1-i)/\sum_{i=1}^{n} i$, $i=1, \ldots, n$. Like *GD-INF* and *GD-FIN*, this pattern implies more rapid capacity depreciation in the earlier years than in the later years of the service life, but the rate of decline in productive capacity is more rapid in this case. This pattern will be denoted as *SYD*.

4) A light bulb or "one-hoss shay" pattern for which no capacity depreciation occurs until the very end of the service life. For this pattern, $d_i=0$ for $i=1, \ldots, n-1$, and $d_n=1$. This pattern will be denoted as *OHS*.

5) A straight-line pattern for which $d_i=1/n$, $i=1, \ldots, n$. According to this pattern, productive capacity declines by the same amount in each year of the service life. It will be referred to in brief as *SL*.

This particular set of capacity depreciation patterns is, of course, arbitrary and by no means exhaustive, although it does represent an interesting spectrum of possi-

---

accounting for physical deterioration, their denominators for obsolescence, and they can readily be converted into a set of d_i, such that

$$1-\sum_{i=0}^{m} d_i = \frac{1-\sum_{i=1}^{m} h_i}{(1+g)^m}$$

From an empirical point of view, we specify a set of d_i to calculate a capital stock (K_t) and net investment (K_t-K_{t-1}), and we need not be concerned with the particular set of h_i and value of g which generate d_i. If embodied technical progress is not occurring, the d_i and h_i are identical; if it is occurring, we presume that our d_i capture it. In the latter case, the production function aggregated over all vintages of capital must include a trend term involving $(1+g)$; but since the trend term will not appear in the marginal productivity conditions for net worth maximization, it is immaterial.

[6] I make the simplifying assumption that capital goods are retired precisely at the end of their assumed service lives. This is unlikely in practice, but information on retirement distributions is largely nonexistent.

bilities. I must emphasize further that although the names and descriptions attached to these depreciation patterns bear similarities to the terminology of economic and tax depreciation accounting, they refer not to economic or tax depreciation, but to capacity depreciation. The relationship between capacity and economic depreciation is taken up below in Section III.

B. *The Demand for Capital*

An important innovation in recent studies of investment behavior has been the inclusion of an implicit rental price of capital as a determinant of the capital stock which firms desire to hold. The rental price measures now being used are based on the particular assumption regarding capacity depreciation discussed in the preceding section, namely, that the pattern of capacity depreciation is characterized by a decaying geometric series. Our first task in this section is to generalize the formulation of the rental price to allow for any pattern of capacity depreciation.

Suppose that firms in a given industry are considering adding one (small) unit to their capital stock.[6] We assume that the price q which they must pay for a new unit of capital, the price p which they receive for each unit of output, and the interest rate r at which they discount future revenues and costs are expected to be constant over the indefinite future. An increase in capital stock K by one unit will increase output X in each future period by $\partial X/\partial K$, the marginal product of capital. Gross revenue in each period will rise by $p(\partial X/\partial K)$, but net revenue will rise by only $p(\partial X/\partial K)-q(\partial R_{jt}/\partial K)$, where $\partial R_{jt}/\partial K$ is the increase in replacement in period t required to maintain the capital stock at its new level, assuming that the jth mortality distribution holds. Firms should add to their capital stock if the present value of the net revenues generated exceeds the price of a unit of capital. Stated algebraically, this condition is:

$$\sum_{t=1}^{\infty} [p(\partial X/\partial K) - q(\partial R_{jt}/\partial K)] \cdot (1+r)^{-t} > q \tag{3}$$

To maximize their net worth, firms should continue to add to their capital stock until (3) holds as an equality. Hence, net worth maximization requires that

$$\frac{\partial X}{\partial K} = \frac{qr\left[1 + \sum_{t=1}^{\infty} (\partial R_{jt}/\partial K)(1+r)^{-t}\right]}{p} = \frac{c_j}{p} \tag{4}$$

The expression c_j is the implicit rental price of capital corresponding to the jth mortality distribution. The rental price consists of two components. The first, rq, may be viewed as the interest per period that firms must pay on funds borrowed to finance an initial net investment of q dollars (one unit of capital). The second, $r\sum_{t=1}^{\infty} q(\partial R_{jt}/\partial K)(1+r)^{-t}$, is the present value of the cost of all future replacements (including replacements of replacements) needed to maintain one unit of capital, multiplied by the discount rate r. It represents, then, the replacement expense per period expressed as an annuity whose present value equals the present value of the cost of all future replacements.

The formula for the rental price can be simplified in the following way. Let us define F_j as the present value of the stream of capacity depreciation on a unit of capital according to the jth mortality distribution; that is,

$$F_j = \sum_{i=1}^{\infty} d_{ji}(1+r)^{-i} \tag{5}$$

---

[6] The presentation here follows my 1968 paper, except that for convenience the existence of a business income tax is ignored. In the empirical work reported on below, tax policy parameters are incorporated in the rental price.

It can be shown that

$$(6)\quad \sum_{t=1}^{\infty} (\partial R_{jt}/\partial K)(1+r)^{-t} = F_j/(1-F_j)$$

so that

$$(7)\quad c_j = qr/(1-F_j)$$

For the special case in which $d_i=\delta(1-\delta)^{i-1}$, which has been assumed in previous studies,

$$(8)\quad F = \sum_{i=1}^{\infty} \delta(1-\delta)^{i-1}(1+r)^{-i} = \delta/(r+\delta)$$

and

$$(9)\quad c = q(r+\delta)$$

which is the rental price expression commonly found in the literature.[7,8]

In this study, rental price estimates for equipment and structures in manufacturing industries are prepared for each of the mortality distributions employed in preparing the net investment series described above. Thus, for each assumed mortality distribution we have net investment and rental price estimates which are consistent with one another.

We have seen that firms should adjust their capital stock to equate the value of the marginal product of capital to its implicit rental price. We shall assume that output is produced according to the Cobb-Douglas production function

$$(10)\quad X = AL^{\alpha}(K_j^e)^{\beta_j}(K_k^s)^{\gamma_k}$$

where X=output, L=labor input, K_j^e =stock of equipment measured in accordance with the jth mortality distribution, K_k^s=stock of structures measured in accordance with the kth mortality distribution, and A, α, β_j, and γ_k are the production function parameters corresponding to the particular measures of the capital stocks. Substituting the marginal-product expressions from (10) into equation (4) and denoting the implicit rental prices of equipment and structures by c^e and c^s, respectively, we obtain the following expressions for the desired, or optimal, capital stocks:

$$(11)\quad (K_j^e)^* = \beta_j \frac{pX}{c_j^e}$$

$$(12)\quad (K_k^s)^* = \gamma_k \frac{pX}{c_k^s}$$

These stock-demand functions, which were originally derived and utilized by Jorgenson (1963), will be used in the investment relations fitted below. Considerable controversy exists over their suitability in investment analysis (see, for example, my 1969 paper), and further work is warranted to determine the sensitivity of results reported below to alternative specifications of the stock-demand functions.

---

[7] While geometric decline in productive capacity has been assumed in the rental price measures used in applied work on investment, there are to my knowledge two theoretical studies that treat other patterns of capacity depreciation and arrive at rental price expressions similar to equation (7). See the works of Kenneth Arrow and Vernon Smith.

[8] When expected prices are not constant, the implicit rental price is in general more complex and less tractable empirically than equation (7). In the special case of geometric decay, the rental price for period t becomes $q_t(r+\delta)-(1-\delta)(q_{t+1}-q_t)$. Presuming that q_t is known to firms, all that we would need to estimate is their expectation of the price of capital goods in t+1, q_{t+1}. However, for other patterns of capacity depreciation, the rental price for period t will depend on the expected prices of capital goods at *every* future date. We could make the simplifying assumptions that 1) firms expect capital goods prices and all other prices to rise at a constant rate over the indefinite future, and 2) they expect the nominal rate of interest each period to reflect fully the rate of inflation, with the real rate of interest remaining constant. If this were so, then equation (7) holds, provided that r is the real rate of interest and F_j is calculated using the real rate of interest. In my empirical implementation of (7), I assume that r is constant over time and that it measures the real marginal rate of return on capital desired by firms. This is consistent either with the assumption that expected prices are constant and equal to current prices or with the set of assumptions stated in the middle of this paragraph.

C. *Investment Demand*

Net investment measures the flow of current new additions to capital stock, and it seems reasonable to relate net investment to current and past changes in the desired capital stock. The lag distribution which I shall assume is the relatively flexible form employed by Hall and Jorgenson in their study of annual investment for total manufacturing. The first two terms of the lag distribution are freely estimated, and the remaining terms are presumed to decline at a geometric rate. The equations for net investment (one for equipment, one for structures) take the form:

$$\begin{aligned}(13)\qquad N_{jt} &= \omega_{1j}(K^*_{jt} - K^*_{j,t-1}) \\ &\quad + \omega_{2j}(K^*_{j,t-1} - K^*_{j,t-2}) \\ &\quad + \lambda_j N_{j,t-1} + u_{jt}\end{aligned}$$

where ω_{1j}, ω_{2j}, and λ_j are the parameters of the lag distribution, and u_{jt} is a random error, everything holding for the jth mortality distribution. For example, if we substitute the expression for the desired stock of equipment into equation (13), we have

$$\begin{aligned}(14)\qquad N^e_{jt} &= b_1\left(\frac{p_tX_t}{c^e_{jt}} - \frac{p_{t-1}X_{t-1}}{c^e_{j,t-1}}\right) \\ &\quad + b_2\left(\frac{p_{t-1}X_{t-1}}{c^e_{j,t-1}} - \frac{p_{t-2}X_{t-2}}{c^e_{j,t-2}}\right) \\ &\quad + b_3N^e_{j,t-1} + u_{jt}\end{aligned}$$

where b_1, b_2, and b_3 are the regression coefficients to be estimated. We see that $b_1=\omega_{1j}\beta_j$, $b_2=\omega_{2j}\beta_j$, and $b_3=\lambda_j$. Employing the proposition that the terms of the lag distribution must sum to unity, we can identify the parameters β_j, ω_{1j}, and ω_{2j}, from estimates of the b coefficients.

Replacement investment is assumed to be equal each year to replacement requirements calculated according to the jth mortality distribution, R_{jt}. Thus, if our net investment relation gives a prediction $\hat{N}_{jt}$ for net investment in year t, according to the jth mortality distribution, then our prediction of gross investment identified with the jth mortality distribution is

$$(15)\qquad \hat{I}_{jt} = \hat{N}_{jt} + R_{jt}$$

Since $I_t = N_{jt} + R_{jt}$, the error in our prediction of gross investment in year t when we assume the jth mortality distribution is

$$(16)\qquad I_t - \hat{I}_{jt} = N_{jt} - \hat{N}_{jt}$$

that is, the prediction error for gross investment is equal to the prediction error for net investment. Thus, while we are actually interested in comparing the ability of alternative mortality distributions to explain gross investment, we are able to make the comparisons by examining the predictive accuracy of the fitted net investment relations.

Although the form of the lag distribution assumed here is relatively flexible, it does have limitations in that it imposes the same lagged response of net investment to changes in all the variables which determine the desired stock of capital. Also, we have assumed that factor proportions can readily be changed both before and after capital goods are put in place—the so-called putty-putty situation. In a putty-clay model, the mortality distribution of capital goods is a crucial determinant of the distributed lag on price changes, since new factor proportions associated with price changes (demand constant) can be incorporated only in replacement of existing capacity.

II. Investment Relations for Equipment and Structures in Manufacturing Industries

A. *The Data*

My empirical analysis of investment behavior covers twenty-one manufacturing industries listed in the Standard Industrial Classification (*SIC*) code. The annual gross investment data used to compute post-World War II measures of net investment and replacement for each mortality distribution were taken from a study prepared by Jack Faucett Associates, giving current and constant (1958) dollar gross investment in equipment for 1921–66 and in structures for 1890–1966. Faucett used the Commerce Department's structures deflator (constant cost 2) for total manufacturing in obtaining constant dollar estimates of structures investment; the same deflator was used in all manufacturing industries. The equipment deflators vary from industry to industry, however, the deflator for each industry being a weighted average of the Commerce Department's deflators for different types of equipment, with the weights reflecting the industry's use of each type of equipment.

Estimates of the implicit rental prices of equipment and structures are based on equation (7) above, except that adjustments are made to account for influences of tax policy. The formula used is:

$$c_j = \frac{q r}{1 - F_j} \cdot \frac{1 - s - \tau(1 - ms)B}{1 - \tau} \tag{17}$$

where s = the rate of investment tax credit, τ = the tax rate on business income, m = the proportion of the investment credit that must be deducted from the depreciable base of assets on which the credit is claimed, and B = the present value of the stream of depreciation for tax purposes stemming from one dollar of current investment. The second fraction on the right-hand side of (17) is the tax adjustment factor, which I have derived in previous studies (1968, 1971). This factor is the same for all mortality distributions. Service lives and depreciation patterns permitted for tax purposes determine B, whereas the mortality distribution determines F. As we experiment with different mortality distributions, we change F but not B.

The Faucett deflators for capital expenditures are used as estimates of the prices of new capital goods, and the discount rate is set at 10 percent per annum, the same for all years and industries. The rate of the investment credit for equipment (there has been no credit for structures investment) was set at 3.5 percent for 1962, 7 percent for 1963–65, and 5.25 percent for 1966. The 1962 and 1966 rates reflect the fact that the credit was available only for parts of these years. The proportion of the investment credit that must be deducted from the depreciable base is unity in 1962–63 and zero thereafter, reflecting repeal of the Long Amendment in the Revenue Act of 1964. The tax rate is set equal to the maximum statutory corporate income tax rate.

The value of B is a weighted average of present values of tax depreciation for three write-off methods: straight-line, double-rate declining-balance (with a switch to straight-line after the midpoint of the asset's life), and sum-of-the-years-digits. For 1947–53, straight-line is given a weight of unity, the others zero. Thereafter, the weights are .293, .354, and .353, respectively, for equipment and .349, .302, and .349 for structures, reflecting the liberalization of depreciation methods which occurred in 1954. These weights are based on information gathered by the Treasury in the 1950's as reported by Norman Ture. The same weights are used for all industries, since more detailed information is not available.

The service life of structures for tax purposes is set at 23 years throughout the 1947–66 period, the same for all industries. This estimate is again based on Ture's examination of Treasury survey data. More detailed information was available for equipment, including average and median service lives for tax purposes at roughly the two-digit *SIC* level for samples of firms in the late 1950's. From these data, judgmental estimates were made for the industries included in this study, and these estimated lives were assumed to hold for the period 1947–61. The guideline lives published by the Treasury in 1962 were assumed to hold for the period 1962–66. The guideline classes do not correspond exactly to the two-digit *SIC* breakdown, so some subjective adjustments were necessary. The final estimates of equipment lives for tax purposes by industry are given below in Table 3 of Section III. The third column of Table 3 gives service life estimates representing the lower bounds of lives permitted under the *Asset Depreciation Range* (*ADR*) *System*, which took effect in 1971. They are not used in estimating the present investment relations, but will be used in Section III where comparisons are made between economic and tax depreciation. Finally, the output measure for each industry is gross product originating.

B. *The Net Investment Regressions*

Net investment relations of the form of equation (14) were fitted to annual equipment and structures expenditures for the period 1949–66.[9] The following procedure was used in selecting the "best" mortality distribution for each type of asset in each industry. The first step involved ranking the regressions for the various mortality distributions in terms of their ability to explain gross investment over the sample period. As we saw above, errors in predicting gross investment for a given mortality distribution are equal to those in predicting net investment for that mortality distribution. Thus, the goodness of fit statistic from the regressions that is appropriate to ranking them is the standard error of estimate. If the mortality distribution yielding the lowest standard error of estimate implied reasonable parameters for the lag distribution and the production function, then it was selected as the best mortality distribution, i.e., the mortality distribution indicated by the behavior of capital expenditures. If implausible parameters for the lag distribution and/or the production function were implied by the best fitting mortality distribution, the search was continued until the best fitting mortality distribution which also had reasonable implications was discovered.

Two criteria were employed in judging the reasonableness of the implied parameters of the lag distribution and the production function. First, the estimate of λ was required to be less than unity. If λ were greater than one, the adjustment of capital stock to its desired level would be unstable. Second, estimates of the elasticities of the stocks of equipment and structures in the production function (β for equipment and γ for structures) were required to be positive and less than unity. Elasticities greater than unity imply increasing marginal products, a violation of the second-order conditions for net worth maximization.

An example will help to illustrate the procedures followed. Consider *SIC* 22, textile mill products, for which the standard errors of estimate are summarized in Table 1. The best fitting equations for equipment were, in descending order, those corresponding to *SL* with a 14-year service life and *SL* with a 12-year service life. The first of these implied an estimate of β of 1.595, while the second implied an estimate of β of .583; hence, the first was rejected and the second selected as best. For structures, there are many equations yielding

---

[9] In most instances, the residuals showed first-order autocorrelation, and the regressions were rerun using transformations of the original variables of the form $y_t - \hat{\rho} y_{t-1}$, where $\hat{\rho}$ is the autocorrelation coefficient computed from the estimated residuals of the initial regression. Since one observation is lost when this transformation is performed, the adjusted equations are fitted to observations for 1950–66, rather than 1949–66. A complete tabulation of the regressions is available on request.

TABLE 1—SUMMARY OF STANDARD ERRORS OF ESTIMATE FROM NET INVESTMENT REGRESSIONS, 1950–66

SIC 22, Textile Mill Products

Assumed Service Life (Years)	Capacity Depreciation Pattern				
	GD-INF	*GD-FIN*	*SYD*	*OHS*	*SL*
A. Equipment					
8	31.21	34.04	36.69	136.3	47.22
10	32.99	27.98	33.57	117.7	36.93
12	33.83	26.22	28.99	127.3	24.03
14	34.46	33.09	26.14	120.2	19.26
16	34.91	59.70	27.50	129.7	37.83
18	35.19	24.65	29.02	69.57	33.24
20	35.36	28.24	31.22	38.00	33.21
22	35.43	29.48	31.19	29.01	29.81
B. Structures					
20	13.48	10.19	13.01	18.09	10.02
25	13.46	10.59	11.75	39.22	8.983
30	13.53	14.69	11.37	17.99	11.51
35	13.56	16.26	11.72	48.83	14.59
40	13.51	15.83	12.65	49.20	17.78
45	13.36	12.77	13.78	21.67	18.79
50	13.12	14.77	14.71	24.01	18.24

better fits than *OHS*-30, which was finally selected, but each one gave an estimate of λ greater than unity.

Examination of Table 1 reveals general characteristics of the results which are worth noting. First, there is considerable variation in the standard errors of estimate among mortality distributions. The empirical model of investment behavior employed here does appear to be sensitive to specification of the mortality distribution and does discriminate among the alternatives, more so in some industries than in others. Second, the standard errors of estimate for a given capacity depreciation pattern often change in a capricious manner as the service life is varied. Changes in the service life can lead to sizable, abrupt changes in the estimates of net investment, particularly when the capacity depreciation pattern is not very smooth. The service life determines the length of the historical period which influences estimates of net investment in the postwar era, and even small changes in the service life may lead to the inclusion or exclusion of rather unusual historical experiences.

Table 2 lists the final mortality distributions selected for each industry and gives the lag distribution and production function parameters implied by the regressions corresponding to those mortality distributions. An interesting feature of Table 2 is the appearance of only one case in which the *GD-INF* capacity depreciation pattern —the one typically used in recent studies of investment behavior—yielded the best results. However, capacity depreciation patterns which imply losses in productive capacity over the service life are predominant for equipment. A very striking finding is that the *OHS* depreciation pattern yields the best results for structures in eleven industries.

In general, the regressions for each industry must be carefully reviewed to determine their reliability. The particular model of investment behavior employed here does not perform equally well in each industry and for each type of asset. More-

TABLE 2—SERVICE LIVES, CAPACITY DEPRECIATION PATTERNS, AND STRUCTURAL PARAMETERS INDICATED BY INVESTMENT REGRESSIONS

SIC Industry Code	Service Life (Years)	Capacity Depreciation Pattern	ω_1	ω_2	λ	β and γ[a]
Equipment						
20	14	*SYD*	.060	.253	.687	.085
21	12	*SL*	.159	−.038	.879	.031
22	12	*SL*	.045	.033	.922	.583
23	14	*SYD*	—	.033	.967	.239
24	12	*SL*	.086	—	.914	.325
25	12	*GD-FIN*	.017	.033	.950	.267
26	18	*OHS*	.003	.032	.965	.591
27	18	*GD-FIN*	—	.030	.970	.437
28	10	*SL*	−.038	.119	.919	.543
29	10	*SL*	−1.314	1.779	.535	.013
30	22	*GD-FIN*	.060	.011	.928	.311
31	10	*GD-FIN*	.428	.166	.406	.016
32	18	*OHS*	−.025	.082	.943	.303
33	10	*SL*	.100	.198	.701	.251
34	10	*GD-FIN*	.141	−.004	.863	.149
35	10	*GD-FIN*	.104	.072	.823	.147
36	18	*GD-FIN*	.186	.177	.636	.097
37+19–371	12	*GD-INF*	.036	−.016	.980	.570
371	8	*GD-FIN*	.077	.363	.559	.076
38	16	*SL*	.353	−.032	.679	.044
39	22	*SL*	.013	−.004	.991	.691
Structures						
20	20	*SL*	−.295	.869	.426	.003
21	30	*OHS*	.022	.010	.969	.009
22	30	*OHS*	.266	.207	.527	.043
23	25	*GD-FIN*	.117	.014	.869	.019
24	20	*SL*	.549	−.310	.761	.013
25	50	*OHS*	.011	.005	.984	.353
26	45	*OHS*	.050	.039	.911	.139
27	25	*SYD*	−.088	.239	.850	.024
28	25	*GD-FIN*	−.086	.131	.954	.117
29	45	*OHS*	.049	.190	.761	.098
30	40	*SL*	.032	.033	.935	.100
31	25	*GD-FIN*	.151	.178	.671	.022
32	50	*OHS*	.033	.121	.847	.132
33	30	*SL*	−.090	.728	.362	.012
34	45	*OHS*	.050	−.012	.962	.088
35	40	*OHS*	.041	.071	.887	.080
36	45	*OHS*	.129	.207	.664	.037
37+19–371	30	*GD-FIN*	1.555	−1.216	.661	.011
371	45	*OHS*	.017	.048	.935	.059
38	50	*OHS*	.073	.144	.783	.057
39	45	*SL*	.001	.008	.991	.589

[a] β for equipment; γ for structures.

over, in many instances a short service life combined with a slow capacity depreciation pattern gave results that were not very different from those obtained for a longer service life combined with a faster capacity depreciation pattern. Perhaps a more discerning test in these cases would be a comparison of postsample predictions from the competing alternatives, but I have not yet attempted this. While the analyses in Section III will treat the mortality distributions in Table 2 as "point estimates," they should more often be viewed as the best (marginally superior) estimates among a set of perhaps three or four candidates suggested by the regressions.[10]

III. Comparisons of Tax Depreciation Practices with Depreciation Indicated by Investment Behavior

A. *Service Lives for Tax Purposes Compared with Estimated Service Lives*

The mortality distributions indicated by investment behavior can be used in a variety of ways in examining depreciation practices for tax purposes. One interesting comparison is that between service lives employed for tax purposes and those revealed by actual capital expenditures. Table 3 provides such comparisons. Equipment service lives for tax purposes prevailing before 1962 were generally above those indicated by investment behavior, in many cases by substantial margins. The 1962 guidelines generally brought tax lives into closer agreement with those revealed by investment behavior; for 10 of the 21 industries, guideline lives were within two years of revealed service lives, compared with only 3 of 21 in the preguideline era. The *ADR* lives currently permitted are below revealed service lives in 15 of the 21 industries, although they are within 2 years of revealed service lives in 10 industries. Only 2 industries, primary metals and motor vehicles, appear to have *ADR* lives substantially above revealed service lives. In both instances, however, Table 3 is very likely misleading. The *ADR* life of 10 years for motor vehicles is for general machinery and equipment. The Treasury has permitted auto producers to depreciate special tools, dies, and jigs over a period of about 3 years, since these items are typically replaced with each major model change. Since they account for a significant portion of equipment expenditures (over 50 percent according to Treasury estimates), the actual tax life in this industry is certainly less than 10 years. The problem in primary metals is the classification of blast furnaces. For tax purposes they are depreciated as equipment, but they are classified as structures in the investment series used here. Hence, the revealed service life in Table 3 probably underestimates the life of those assets actually depreciated as equipment.

According to the weighted-average equipment lives at the bottom of the table, 1) the preguideline average tax life for manufacturing equipment exceeded the revealed service life by 3.6 years, or about 28 percent, 2) the guidelines equalized the average tax life and the revealed service life, and 3) *ADR* reduced the average tax life to a level about 19 percent below the revealed service life. While these averages provide additional perspectives on the findings, they can and do, of course, conceal serious inequities or advantages ap-

---

[10] Tests of the significance of differences in the variances of residuals obtained for alternative mortality distributions might seem appealing, but they would be quite complex and, I believe, suspect. Given the research design, there is good reason to believe that residuals for any two mortality distributions will be highly correlated, and in fact they generally are. Hence, the conventional F-test for equality of two variances, which is founded on the assumption that the variables are independently distributed, is not appropriate. Also experimentation with such a large number of mortality distributions has certainly exhausted the degrees of freedom in this sample.

TABLE 3—SERVICE LIVES FOR TAX PURPOSES COMPARED WITH THOSE INDICATED BY INVESTMENT BEHAVIOR
(Years)

SIC Industry Code	Equipment Life for Tax Purposes			Excess of Tax Life Over Life Indicated by Investment Behavior Equipment			Structures All Years
	1947–61	1962–70	*ADR*[b]	1947–61	1962–70	*ADR*	
20	18	15	12	4	1	−2	3
21	18	15	12	6	3	0	−7
22	16	12	10	4	0	−2	−7
23	13	9	7	−1	−5	−7	−2
24	15	10	8	3	−2	−4	3
25	15	10	8	3	−2	−4	−27
26	19	14	11	1	−4	−7	−22
27	15	11	9	−3	−7	−9	−2
28	13	11	9	3	1	−1	−2
29	18	14	11	8	4	1	−22
30	15	13	10	−7	−9	−12	−17
31	14	11	9	4	1	−1	−2
32	23	17	14	5	−1	−4	−27
33	20	16	13	10	6	3	−7
34	16	12	10	6	2	0	−22
35	15	12	10	5	2	0	−17
36	15	11	9	−3	−7	−9	−22
37+19–371	16	12	10	4	0	−2	−7
371	15	12	10	7	4	2	−22
38	16	12	10	0	−4	−6	−27
39	14	12	10	−8	−10	−12	−22
All Industries[a]	16.5	12.9	10.5	3.6	0.0	−2.4	−12.4

[a] Weighted average, with weights proportional to equipment or structures expenditures in 1966.
[b] *Asset Depreciation Range (ADR) System.*

parently experienced by particular industries.

Table 3 indicates that revealed lives for structures are generally much longer than the 23-year average tax life in manufacturing. Indeed, in 11 industries revealed lives range from about 74 percent to 117 percent above tax lives. The weighted averages at the bottom of the table show a tax life for manufacturing structures that is 12.4 years, or about 35 percent, below the revealed life. While the long lives indicated by investment behavior in several industries are far out of line with the average tax life in manufacturing, they are not out of line with the service lives of 45 to 50 years set forth in the Treasury's *Bulletin F* and in the 1962 *Guidelines*. Indeed, in light of the Treasury's prescribed lives, the observed average tax life of 23 years seems surprisingly low.

B. *Comparisons of Estimated Economic Depreciation and Tax Depreciation*

Even though an industry may be using a tax life that corresponds to its retirement practices, it may be writing off assets over the tax life at a rate that exceeds or falls short of actual economic depreciation. By comparing the present values of the streams of tax and economic depreciation on one dollar of capital expenditures, we can gain additional insight into the impact of depreciation policy. To do this, we must first see how our mortality distributions which relate to capacity depreciation can be translated into patterns of economic depreciation.

If prices are expected to remain constant or to change at a constant rate with r being the real rate of interest, then the value of an asset at the beginning of year t of its service life V_t should equal the present value of the stream of productive capacity from year t to the end of the service life (year n). For example, the *OHS* capacity depreciation pattern specifies that productive capacity is the same in each year of the service life; assume that it is unity. Then

$$V_t = \sum_{i=0}^{n-t} (1+r)^{-i} \tag{18}$$
$$= (1+r)[1-(1+r)^{t-n-1}]/r$$

Economic depreciation in year t is the loss of value during the year, or $V_t - V_{t+1}$; and $v_t = (V_t - V_{t+1})/V_1$ measures the loss in value relative to the value when new. Hence, the v_t describe the pattern of economic depreciation on an asset whose initial value is one dollar, and $\sum_{t=1}^{n} v_t(1+r)^{-t}$ is the present value of the stream of economic depreciation. For reasonable discount rates, the *OHS* pattern implies that economic depreciation occurs at a slower pace in the early years of an asset's life than in the later years, but the other patterns imply the reverse. And in the case of *GD-INF*, we have the special result that economic and capacity depreciation are identical in each year.

Table 4 presents ratios of present values of tax to estimated economic depreciation. Ratios for equipment are shown for four eras of depreciation policy. Tax allowances in the preguideline, pre-acceleration era (1947–53) were well below economic de-

TABLE 4—PRESENT VALUES OF TAX DEPRECIATION AS PERCENTAGES OF PRESENT VALUES OF ECONOMIC DEPRECIATION ON ONE DOLLAR OF CAPITAL EXPENDITURES

SIC Industry Code	Equipment				Structures	
	1947–53	1954–61	1962–70	ADR	1947–53	1954–69
20	67.4	78.0	84.9	92.9	76.3	90.5
21	70.8	82.0	89.3	97.7	232.5	275.9
22	75.9	86.6	97.7	104.0	232.5	275.9
23	80.6	90.1	102.2	109.3	82.0	97.2
24	78.7	89.3	104.0	111.2	76.3	90.5
25	76.2	86.5	100.8	107.7	989.7	1174.5
26	122.6	143.2	164.9	180.8	677.2	803.5
27	90.2	102.3	115.5	123.1	71.5	84.8
28	79.7	89.1	94.7	101.0	82.0	97.2
29	66.6	77.1	86.4	94.7	677.2	803.5
30	100.0	113.4	120.3	132.1	130.0	154.2
31	74.5	83.9	91.9	98.0	82.0	97.2
32	107.5	129.0	151.3	164.9	989.7	1174.4
33	62.2	73.0	81.5	89.1	101.0	119.9
34	69.3	79.0	89.1	94.9	677.2	803.5
35	71.8	81.4	89.1	94.9	470.7	558.5
36	90.2	102.3	115.5	123.1	677.2	803.5
37+19–371	78.2	89.3	100.6	107.2	91.9	109.0
371	67.6	76.7	83.9	89.3	677.2	803.5
38	85.8	97.9	110.4	117.5	989.7	1174.4
39	110.3	124.1	131.9	140.5	130.0	154.2
All Industries[a]	80.5	92.3	102.8	110.8	374.9	444.9

[a] See Table 3.

preciation in all but 4 industries. While service lives for tax purposes in 2 of these industries exceeded the service lives indicated by investment behavior, even the straight-line write-off method gave rise to tax depreciation that was considerably more rapid than that implied by the capacity depreciation patterns evidenced by equipment purchases in these industries. For many industries in which the pre-guideline lives were roughly in accord with revealed service lives, the present value of tax depreciation fell short of economic depreciation by 15 to 20 percent, because the straight-line write-off method was too slow relative to the implied patterns of economic depreciation.

The adoption of accelerated write-off methods in the second era generally increased the present values of tax depreciation relative to economic depreciation by about 10 percentage points. This was sufficient to add 3 industries to the list of those for which the present value of tax depreciation roughly equalled or exceeded that of economic depreciation. The guideline lives identified with the third era further increased the present values of tax depreciation, though not as uniformly as did the accelerated write-off methods. Under the guidelines, 12 of the 21 industries were permitted tax write-offs that roughly coincided with or exceeded economic depreciation.

The *ADR* system instituted in 1971 increased the present values of tax depreciation by about 6 to 10 percentage points relative to economic depreciation. Since *ADR* involved a 20 percent reduction in tax lives across the board, the largest benefits were conferred on industries whose pre-*ADR* tax lives were longest. If these were the industries experiencing the least favorable treatment prior to *ADR*, then *ADR* would have moderated existing inequities; but close inspection of Tables 3 and 4 reveals that this was not entirely the case. The weighted averages given in the last row of the table indicate that under the guidelines the present value of tax depreciation exceeded economic depreciation of all manufacturing equipment by 2.8 percent; under *ADR*, this figure rose to 10.8 percent.

Service lives of structures for tax purposes have not been altered since 1947 (at least not by statute or administrative ruling), so that only the pre- and post-acceleration eras need be distinguished. The picture revealed by Table 4 is a very mixed one indeed. Since service lives indicated by investment behavior are in many instances a decade or two longer than the 23-year average tax life, it is not unusual for the present values of tax depreciation to exceed economic depreciation by substantial amounts. This tendency is further strengthened by the relatively slow capacity depreciation of structures evidenced by investment behavior in many industries. Even if tax lives permitted in many of these industries were increased to the 40–50-year lives indicated by investment behavior, tax write-offs would still be excessive due to the use of uniform or accelerated depreciation for tax purposes, when economic depreciation occurs at the decelerated pace associated with *OHS* capacity depreciation.

IV. Conclusions

The new approach to determination of service lives and capacity depreciation patterns of industrial assets pursued in this study appears to be a promising one. The present findings pertaining to manufacturing industries may be summarized as follows:

1) Tax service lives for equipment prescribed by the Treasury in 1962 were on the average in line with service lives revealed by investment behavior; consequently, the lower-bound *ADR* lives currently permitted are roughly 20 percent

shorter than revealed service lives. Revealed service lives of structures are substantially longer than those used for tax purposes in the postwar period, although they are on the average somewhat shorter than the lives prescribed by the Treasury in *Bulletin F* and the 1962 *Guidelines.*

2) Geometric decay of productive capacity—a commonly employed assumption in recent studies of investment behavior—does not appear to underlie actual capital spending decisions. Equipment generally evidences losses in productive capacity as it ages, though not necessarily at a geometric rate, but structures in the majority of industries suffer no loss in productive capacity over their service lives (they resemble one-hoss shays).

3) The derived estimates of economic depreciation patterns indicate that write-offs permitted under current tax regulations are on the whole generous, particularly for structures. While equipment write-offs under the 1962 *Guidelines* were on the average approximately in line with estimated economic depreciation, the *ADR* system instituted in 1971 allows excessively rapid tax depreciation.

It should be recalled that these are the *general* findings; they do not necessarily hold for each of the 21 industries studied. Of course, the present results must be viewed cautiously, since they are based on a particular, highly constrained model of investment behavior and on statistical criteria that do not include evaluations of the postsample predictive performance of alternative mortality distributions. There is some danger that the mortality distributions selected here just happen to yield good results over the period of fit, and in some instances the sample period fits were nearly identical for several mortality distributions. Also, it has been assumed throughout this work that actual service lives were constant over the period studied—not just the sample period for the investment regressions, but the entire historical period over which gross investment outlays were cumulated to obtain postwar estimates of net investment and replacement. Service lives are, however, decision variables for firms, and their optimal values may be influenced by tax depreciation policy, income tax rates, and investment credits, among other things.[11] This complex issue merits careful attention, but the theory of optimal service lives has not yet been shaped in an empirically operational form for studying investment behavior.

---

[11] See sections 5 and 6 of the paper by Martin Feldstein and Michael Rothschild.

REFERENCES

K. J. Arrow, "Optimal Capital Policy, the Cost of Capital, and Myopic Decision Rules," in *Annals Inst. Statist. Mathematics,* Tokyo 1964, *16,* 21–30.

P. Cagan, "Measuring Quality Changes and the Purchasing Power of Money: An Exploratory Study of Automobiles," *Nat. Banking Rev.,* Dec. 1965, *3,* 217–36.

R. M. Coen, "The Effect of Cash Flow on the Speed of Adjustment," in G. Fromm, ed., *Tax Incentives and Capital Spending,* Washington 1971, 131–96.

———, "Effects of Tax Policy on Investment in Manufacturing," *Amer. Econ. Rev. Proc.,* May 1968, *58,* 200–11.

———, "Tax Policy and Investment Behavior: Comment," *Amer. Econ. Rev.,* June 1969, *59,* 370–79.

R. Eisner, "Components of Capital Expenditures: Replacement and Modernization Versus Expansion," *Rev. Econ. Statist.,* Aug. 1972, *54,* 297–305.

M. S. Feldstein and D. K. Foot, "The Other Half of Gross Investment: Replacement and Modernization Expenditures," *Rev. Econ. Statist.,* Feb. 1971, *53,* 49–58.

——— and M. Rothschild, "Towards an Economic Theory of Replacement," disc. pap. 249, Harvard Inst. Econ. Res., Aug. 1972.

Z. Griliches, "The Demand for a Durable Input: Farm Tractors in the United States, 1921–57," in A. C. Harberger, ed., *The Demand for Durable Goods,* Chicago 1960, 181–207.

R. E. Hall, "The Measurement of Quality Change from Vintage Price Data," in Z. Griliches, ed., *Price Indexes and Quality Change,* Cambridge 1971, 240–71.

——— and D. W. Jorgenson, "Tax Policy and Investment Behavior," *Amer. Econ. Rev.,* June 1967, *57,* 391–414.

D. W. Jorgenson, "Capital Theory and Investment Behavior," *Amer. Econ. Rev. Proc.,* May 1963, *53,* 247–59.

———, "The Economic Theory of Replacement and Depreciation," disc. pap. 203, Harvard Inst. Econ. Res., Sept. 1971.

A. I. Mendelowitz, "The Measurement of Economic Depreciation," in *1970 Proc. Amer. Statist. Assn., Bus. Econ. Statist. Sec.,* Washington 1971, 140–48.

W. Ramm, "Measuring the Services of Household Durables: The Case of Automobiles," in *1970 Proc. Amer. Statist. Assn., Bus. Econ. Statist. Sec.,* Washington 1971, 149–58.

P. A. Samuelson, "Tax Deductability of Economic Depreciation to Insure Invariant Valuations," *J. Polit. Econ.,* Dec. 1964, *72,* 604–06.

V. L. Smith, *Investment and Production: A Study in the Theory of Capital-Using Enterprise,* Cambridge 1961.

R. M. Solow, "Investment and Technical Change," in K. J. Arrow et al., eds., *Mathematical Methods in the Social Sciences,* Stanford 1960, 89–104.

P. Taubman and R. Rasche, "Economic and Tax Depreciation of Office Buildings," *Nat. Tax J.,* Sept. 1969, *22,* 334–45.

N. B. Ture, *Accelerated Depreciation in the United States, 1954–60,* New York 1967.

F. Wykoff, "Capital Depreciation in the Postwar Period: Automobiles," *Rev. Econ. Statist.,* May 1970, *52,* 168–72.

Jack Faucett Associates, *Development of Capital Stock Series by Industry Sector,* TR-81, Washington, Mar. 1973.

U.S. Department of Commerce, *Fixed Nonresidential Business Capital in the United States, 1925–1970,* Nat. Tech. Info. Serv. pub. no. COM-71-0111, Springfield, Va. 1971.

U.S. Department of the Treasury, Bureau of Internal Revenue, *Income Tax Depreciation and Obsolescence: Estimated Useful Lives and Depreciation Rates, Bulletin F,* rev., Washington 1942.

———, Internal Revenue Service, *Depreciation Guidelines and Rules,* pub. no. 456 (7–62), Washington 1962.

———, *Asset Depreciation Range (ADR) System,* Washington 1971.

Wealth Neutrality and Local Choice in Public Education

By Martin S. Feldstein*

A series of recent judicial decisions has focused public attention on the issue of local choice in the provision of public education. In *Serrano vs. Priest*, *Rodriguez vs. San Antonio* (1971), and similar cases in other states, the lower courts confirmed that education is a responsibility of the state government and held that local expenditures on education may not be a function of the taxable wealth of the local community.[1] Although the United States Supreme Court has overturned these decisions in the appeal of *Rodriguez vs. San Antonio* (1973), the pressure to change the current system remains strong. The Supreme Court majority indicated that its decision reflected the limits of the federal constitutional authority and was not an approval of the status quo in educational finance. Litigation is now likely to shift to challenging the current methods as unconstitutional under state constitutions which, unlike the federal constitution, do deal specifically with education.[2] Moreover, fundamental changes in the financing of local education may not require further pressure from the courts; state legislatures may seek to neutralize the effects of local wealth differences even if the current systems are not held to be unconstitutional.

These judicial decisions and the ensuing legislative proposals run counter to the general economic view of local government finance. The basic presumption of economic analysis is that, because local governments can select different levels of service and because individuals can choose their area of residence, decentralized finance by local governments allows the provision of public services to reflect the variety of individual preferences for public services.[3] Although the level of local spending may be nonoptimal because of intercommunity externalities and because of the method of local budget determination, fiscal decentralization still remains the only alternative to the insuperable problem of determining the optimal level of expenditure on a public service provided by a central government. In effect, autonomous decentralized financing of education provides a quasi market in which households can exercise their diverse preferences by their location decisions.

This paper considers the problem of

---

* Professor of economics, Harvard University. I am grateful to Charles Clotfelter for assistance with the statistical analysis, to Stephen Weiss for providing unpublished data on school expenditures, and to the Ford Foundation and National Science Foundation for financial support. I have benefited from discussions of an earlier version in seminars at Harvard, M.I.T., and Berkeley, and from comments by Noel Edelson, Eric Toder, and David Stern. An earlier and more complete discussion of this study was distributed as Harvard Institute of Economic Research paper no. 293, May 1973 (revised July 1973).

[1] In *Serrano vs. Priest*, the landmark case in this area, the plaintiff and the courts were very much influenced by the line of argument and suggested remedies developed in John Coons et al. For a further discussion of the legal precedents, see Arthur Wise.

[2] Almost immediately after the United States Supreme Court decision in *Rodriguez vs. San Antonio*, the New Jersey Supreme Court held that the current system of local finance violated the New Jersey state constitution. See Wise for a summary of the provisions of other state constitutions.

[3] Charles Tiebout presented a formal analysis of the full efficiency of local government provision of public services under quite special conditions. See Wallace Oates and James Buchanan and Charles Goetz for a further discussion of these issues.

financing local public education in a way that partially or completely neutralizes the effects of local wealth differences without sacrificing the opportunity for local choice. The issues dealt with are likely to be important not only in educational finance but also in other areas of intergovernmental fiscal relations. Moreover, although the entire discussion in this paper is in terms of intergovernmental grants, the same principles can be applied in the design of subsidies to individuals. I return to this in the concluding section.

I. Achieving Wealth Neutrality

In the past few years, there has been increasing attention to the task of combining equity and efficiency in the solution of public sector problems. Specific studies have dealt with optimal taxation, public sector pricing, and the evaluation of public expenditure decisions. In these studies, the consumption of goods and leisure affect social welfare through individualistic utility functions, i.e., the preferences of each individual are assumed to "count." Distributional considerations enter because of the assumed concavity of the individual utility function or of the social welfare function.

A quite different type of distributional concern, which might be called "categorical equity," has been prominent in public debate and legal analysis. The principle of categorical equity singles out particular categories of services such as education and health care which are deemed to be "fundamental interests" and asserts that individuals' consumption of these services should not be allowed to differ substantially or, alternatively, that such differences should not bear a strong relation to the individuals' ability to pay.[4] The current study is limited to the second form of categorical equity, which I shall refer to as "wealth neutrality."

The general problems of achieving wealth neutrality in a decentralized system and of balancing categorical equity and other aspects of economic welfare are well illustrated by the issues of educational finance. In the basic *Serrano vs. Priest* decision, the California Supreme Court asserted the principle of wealth neutrality by stating that the California system of education is unconstitutional because "it makes the quality of a child's education a function of the wealth of his parents and neighbors" (p. 504). The court also emphasized that its decision did not require the state to spend the same amount per pupil in all districts and did not imply that the use of the local property tax is per se unconstitutional. The court did not propose any specific remedies but ordered the state to devise a new method of educational finance.

Two alternative methods are now being considered by state authorities and educational finance specialists: full state assumption of all school finance and the so-called "power equalizing" formula for state aid to local communities (see John Coons, William Clune, and Stephen Sugarman). The full state assumption of school finance would in fact equalize per pupil spending in all districts and would completely eliminate local financial decision making in education.[5] Since the purpose of the current paper is to consider ways of achieving wealth-neutrality without sacrificing local choice, the proposal for full state assumption of educational finance will not be considered. The power equalizing formula will be examined below and criticized for its almost certain failure to achieve its goal of complete wealth neutrality. An alternative "wealth neutralizing matching grant" will be described which has the virtue of preserving local choice while permitting state governments both to achieve any desired degree of wealth neutrality and to provide any desired stimulation to the average level of school spending.

---

[4] Note that the principle of categorical equity is therefore quite different from the traditional notions of vertical equity and horizontal equity (see Richard Musgrave), in which all goods and services are treated alike. Although Musgrave's concept of "merit wants" does single out particular categories of goods, the concern is not with the *distribution* of consumption of these goods but rather with encouraging *additional* consumption of goods that are deemed to be more "meritorious" than individual consumption decisions recognize. While a merit good may also be the subject of categorical equity, it is possible to want to increase the equality of consumption without increasing and possibly while decreasing the average level. In contrast, Kenneth Arrow's recent paper deals with a utilitarian form of categorical equity.

[5] Some proponents of the state assumption of educational finance advocate equalizing local spending per

It is useful to restate the *Serrano vs. Priest* mandate that educational spending not be a function of local wealth as the more precise rule that the elasticity of local educational spending per pupil with respect to the value of local taxable property per pupil be zero. The court's decision in the case of *Serrano vs. Priest* makes it clear that the "quality" of education is to be interpreted as expenditure per pupil and that local "wealth" is to be interpreted as the fair market value of taxable property. Alternative (and probably more appropriate) measures of household wealth could be used in the same way. Similarly, expenditure per pupil could be modified to reflect local differences in input prices or student abilities.

More generally, the wealth elasticity of school spending can serve as a measure of the degree of wealth neutrality or nonneutrality that has been achieved. The constant elasticity log-linear relation between expenditure per pupil in community i (E_i) and wealth per pupil (W_i) may be written:

$$(1) \qquad \ln E_i = \alpha_0 + \alpha_1 \ln W_i + u_i$$

where α_0 is a constant and u_i is a residual that is uncorrelated with $\ln W_i$. The parameter α_1 measures wealth neutrality; complete wealth neutrality requires $\alpha_1 = 0$. Note that while complete equality would imply that the variance of u_i is also zero, this is not required for complete wealth neutrality.

Consider now a more general constant elasticity equation describing the relation of expenditure to wealth, to the net price that the community pays per dollar of educational expenditure P_i, and to other variables X_{ji} that need not be defined at this point:

$$(2) \qquad \ln E_i = \beta_0 + \beta_1 \ln W_i + \beta_2 \ln P_i + \sum_{j=3}^{K} \beta_j \ln X_{ji} + \epsilon_i$$

The price variable P_i reflects the extent and form of grants from the state government to the locality; if m_i is the amount of additional state aid received by the town when it spends an additional dollar on education (i.e., the effective matching rate), the net price is defined by $P_i = 1 - m_i$. Equation (2) implies that the state government can neutralize the effect of intercommunity wealth differences by using a matching grant formula in which the rate of matching varies inversely with local wealth. More specifically, the wealth neutrality coefficient (α_1 of equation (1)) is related to the behavioral parameters of equation (2) by:

$$(3) \qquad \alpha_1 = \beta_1 + \beta_2 \cdot \gamma_{pw} + \sum_{j=3}^{K} \beta_j \gamma_{x_j w}$$

where γ_{pw} is the constant elasticity of P with respect to W and $\gamma_{x_j w}$ is the elasticity of X_j with respect to W. This relation holds both for the "true" population parameters (α, β, and γ) and also for the least

---

"equivalent pupil" and would allow local school districts to supplement state funds by up to 10 percent. Full state assumption has been supported by the Advisory Commission on Intergovernmental Relations, the President's Commission on School Finance, the New York State Commission, and others.

squares estimates that would be used to make the analysis operational.[6]

It is convenient to combine the partial wealth elasticity (β_1) and the other terms that do not relate to price into a single "adjusted" wealth elasticity ($\beta_w = \beta_1 + \sum_j \beta_j \gamma_{x_j w}$) and to relabel the partial price elasticity ($\beta_p = \beta_2$). Equation (3) may now be rewritten:

$$\alpha_1 = \beta_w + \beta_p \gamma_{pw} \tag{4}$$

This is the key equation in analyzing the wealth neutralizing effects of government grants. Since β_w and β_p are behavioral parameters of the educational system, equation (4) implies that the state government can achieve any desired degree of wealth neutrality by a system of matching grants that has the corresponding elasticity of price with respect to wealth (γ_{pw}).

More specifically, the price formula:

$$p_i = kW_i^{\gamma_{pw}} \tag{5}$$

will achieve the wealth neutrality implied by equation (4). Complete wealth neutrality ($\alpha_1 = 0$) is obtained simply by setting $\gamma_{pw} = -\beta_w/\beta_p$. Note that the parameter k of the price formula has the effect of increasing (or decreasing) the level of spending in all towns by the factor k^{β_p}. The state can therefore achieve any desired stimulation of local spending as well as any desired degree of wealth neutrality. Alternatively, the state can achieve the desired wealth neutrality subject to a constraint on the total state budget. By analogy with Jan Tinbergen's analysis of macro-economic policy, the state can achieve any two "targets" (the degree of wealth neutrality and the average level of E or the budget constraint) because it chooses two policy values (k and γ_{pw}).

The wealth neutrality relation of equation (4) shows the inappropriateness of the "district power equalizing" (*DPE*) approach that has been so widely advocated as a method of achieving complete wealth neutrality. Coons et al. use the term district power equalizing to describe a state's system of educational finance if any two school districts that levy the same rate of tax on local wealth have the same level of educational spending. In a sense, *DPE* gives each town the same tax base per pupil; district "power" equalizing is really district "tax base" equalizing. The basic *DPE* formula is equivalent to a percentage equalization grant; i.e., in the current notation the matching rate is $m_i = 1 - kW_i$. The price is thus simply proportional to wealth: $P_i = kW_i$ and therefore $\gamma_{pw} = 1$. It is clear from equation (4) that *DPE* would only be completely wealth neutralizing ($\alpha_1 = 0$) in the very special case in which the absolute wealth and price elasticities were equal.

More generally, with *DPE*, $\alpha_1 = \beta_w + \beta_p$. If the adjusted wealth elasticity exceeds the price elasticity ($\beta_w > -\beta_p$), *DPE* would still leave a positive elasticity of educational expenditure with respect to wealth. Conversely, if the price elasticity exceeds the wealth elasticity, *DPE* would invert the usual wealth-spending relation. The advocates of *DPE* essentially ignore the fact that it may leave a substantial relation (positive or negative) between wealth and school spending because they misinterpret what *DPE* actually does. It is argued that *DPE* " . . . eliminates wealth as a factor determining the offering" (Coons et al., p. 104) of education because they regard *DPE* as changing the wealth of the town when it actually changes the town's net price. A related confusion is treating the local tax *rate* as a price or a measure of "effort." Although *DPE* is sometimes interpreted as equalizing the "price" per dollar of educational services (where price is the rate of tax on local wealth), the relevant price is the local revenue per dollar of educational services purchased and not the rate of tax. Application of *DPE* does not equalize the price of educational services. Similarly, the local tax rate is not a measure of effort unless it is assumed that the "sacrifice" per dollar of revenue varies in proportion to the inverse of local wealth, a utilitarian argument with a Bernoullian utility function. It is this confusion about the nature of the price and wealth effects that leads Coons et al. to write that "because [with *DPE*] incentive is stripped of differential wealth obstacles, it is hoped that the distribution of high and low expenditure districts will become random as regards wealth" (p. 203). Only if the price elasticity and the adjusted wealth elasticity are exactly offsetting would *DPE* achieve complete wealth neutrality.[7]

---

[6] Equation (3) is a special case of Henri Theil's result on specification bias in least squares regression. Since the regressions can be calculated using observations on all towns in the state, the coefficients may be regarded as *measures* of the descriptive statistics rather than as *estimates* of unobservable parameters. The choice of interpretation is irrelevant as long as the operational measure of wealth neutrality is the calculated wealth elasticity of equation (1).

The analysis of this section has shown that the relevant aspects of the local demand for education can be summarized by two key parameters: the price elasticity β_p and the adjusted wealth elasticity β_w. A knowledge of these parameters is required to design a system of matching grants that achieves any desired degree of wealth neutrality and, more generally, to assess the impact on wealth neutrality of any system of matching grants. I turn now to the problems of estimating these parameters.

II. Estimating the Effects of Wealth and Intergovernmental Grants

Although there have been many studies of the expenditures of state and local governments, there are no satisfactory estimates of the effects of intergovernmental grants on local government spending. The essential feature of an "open end" matching grant is that it lowers the net price paid by the local government. In contrast, a block grant has no price effect (unless the community would otherwise have spent less than the amount of the grant). The "closed end" matching grant, i.e., a matching grant with a maximum intergovernmental transfer, has aspects of both the open end matching and block grants. More specifically, for a community that does not reach the maximum, the upper limit is irrelevant and the grant is equivalent to an open end matching grant. For a community that exceeds the maximum, the grant has no marginal price effect and acts instead as a block grant.

In the empirical analysis of local government spending it is therefore crucial to distinguish grants that are effectively matching grants from grants that are effectively block grants. Moreover, the matching grant should be described in terms of the price that it implies for each local community. Unfortunately, previous empirical studies of local government spending have generally ignored the distinction between block grants and matching grants. They therefore described all aid by the volume of funds transferred and did not isolate a price effect.[8] The only studies that distinguished matching grants and used the matching rate as a price variable (see Robert Inman (1970, 1971), David Stern) have constrained the specification of the expenditure equation in a way that precludes an independent estimate of the price elasticity.

---

[7] Coons et al. also discuss a more general form of *DPE* plan in which $P_i=f(t)W_i$ where $f(t)$ is any function of the rate of tax t but a function that is independent of W_i. Although the choice of an $f(t)$ function would obviously affect the resulting wealth neutrality, it might not be possible to design an $f(t)$ that achieves complete wealth neutrality. Analysis of the effects of such a plan is more complicated than with the basic percentage equalization form of *DPE* because $P_i=f(t_i)W_i$ implies that the town does not face a fixed price but rather a price that varies as it changes the quantity of education that it wants to purchase, i.e., as it changes t. The important point here is that the proposed *DPE* plans are unlikely to achieve a satisfactory degree of wealth neutrality because they have not been designed with that in mind.

[8] This is true of the studies of educational spending by Edward Renshaw, George Bishop, and Jack Osman as well as more general studies of government spending by Seymour Sacks and Robert Harris, Roy Bahl and Robert Saunders, James Henderson, Edward Gramlich (1968, 1969), and others.

The current study relates educational expenditure per pupil in Massachusetts towns to the price implied by the matching rate, the local property value per pupil, and other potential determinants of local school spending. The data for Massachusetts are particularly suitable for estimating the effect of matching grants. The matching rate varies substantially among the local communities; in the sample of 105 towns used in the current study, the matching rate ranged from zero to 0.38. This implies that the local price per dollar of educational spending varies between $0.62 and $1.00.

It is important to examine the source of this variation in local matching rates. Clearly, if the state sets the matching rate as a function of variables that are themselves determinants of local spending, the expenditure equation may be underidentified. For example, if the local price were exactly proportional to the local wealth variable, it would be impossible to estimate separate coefficients for price and wealth in a linear or log-linear regression. If the price variable were some non-linear function of wealth, separate coefficients for price and wealth could be estimated in a linear equation only because the analysis introduces the implicit a priori information that the expenditure relation is linear. The justification for such an a priori restriction might be very weak. The linear or log-linear specification, which is generally used only as a convenient approximation, would in this situation be the source of the crucial identifying restriction. Stated somewhat differently, if the legislature uses a non-linear matching grant because it perceives a non-linear effect of wealth on local spending, why should the econometric analysis ignore this information and reject the legislature's perception?

Posing the problem in this way suggests an answer. The actual development of formula matching grants reflects history, legislative compromise, and accident rather than empirical analysis and economic logic. If the matching rate varies sufficiently and is not highly correlated with the other explanatory variables, the possibility of underidentification is not likely to be a problem of practical significance. The Massachusetts experience is a good illustration. Although the basic formula makes state matching proportional to local wealth, a variety of limits and special provisions so attenuates this relation that the correlation between price and wealth in 1970 was only 0.38.[9]

In all of the equations presented below, the dependent variable is current educational expenditures per pupil E, including local, state, and federal funds. The wealth variable W is the official "equalized" or "true" property value per public school pupil, and the price variable P is equal to one minus the local marginal matching rate.[10] Current income INC is measured as median family income in the town. The values of state block grants *SBG* and of federal grants *FG* are also included.[11] Since the federal grants are not block grants but are direct funding of special programs in eligible communities, their effect on spending should be more substantial than that of the state block grants.[12]

---

[9] Legislative compromises modified the basic proportionality formula by imposing both upper and lower limits on the matching rate and on the levels of expenditure that would be matched. The 1966 law also provided that towns would not receive less in the future than they had in a previous base year.

[10] For those towns for which the state aid was effectively a block grant, i.e., for which an additional dollar of local spending would induce no additional state aid, the price is one. For all other towns, the state aid is a proportional matching grant and the price is calculated as the ratio of local current "reimbursible expenditures" (as defined by ch. 70 of the Massachusetts laws) to the sum of that expenditure and the state matching aid implied by the legislative formula. This implicitly assumes that the market prices of educational goods and services are the same in all towns within the state. While this is quite reasonable for the manufactured goods used by the schools, the assumption is more doubtful for teachers' salaries and some other locally purchased services. While the salary of teachers varies

If the local voters assume that they bear less than the full cost of taxes raised from nonresidential property, those towns in which the residential property proportion of the local tax base *RES* is relatively high should prefer a lower level of school spending. If local residents assume that they bear none of the local tax on nonresidential property, the price of educational services is effectively lowered by a fraction equal to the nonresidential property proportion of the local tax base. Local residents will however bear the tax on nonresidential property to the extent that it affects local wages and prices and, more importantly, through its long-run effects on the size of the nonresidential property tax base.[13]

Towns in which the use of private schools is more common may also tend to offer a lower level of school spending although this is offset to the extent that parents support high quality public primary education and then send their children to private secondary schools or support high quality public secondary education but use private (including parochial) schools for the lower grades. The number of private school pupils per capita *PRIV* is included to assess the net effect of these forces. A greater number of public school pupils per capita *PUP* will generally result in less taxable property per pupil (a lower value of *W*). It may also reduce spending by lowering family income per pupil or, alternatively, raise spending by increasing the fraction of families with school age children. The final variable included in the equation is a measure of the recent growth in the number of pupils; more specifically, *GROW* is the ratio of the number of pupils to the corresponding number five years earlier. If there is a reluctance to raise tax rates quickly, a rapid growth of pupils will temporarily reduce per pupil spending. To allow for a more general lagged response of educational spending to all of the explanatory variables, a proportional adjustment model is examined by including the lagged value of the dependent variable among the regressors.

III. The Parameter Estimates

The estimated equations are specified as linear in the logarithms of the variables. The data are a sample of 105 Massachusetts towns that contain 72 percent of the state population.[14] When values for more than one year are used in the same equation, all dollar amounts are converted to 1970 dollars.[15]

---

among towns, it is not clear how much of this variation reflects quality differences. In some estimates not reported in the paper, an educational input price index was included and found to be insignificant and sometimes to have the "wrong" sign; this suggests that much of the price variation reflects quality differences and therefore local choice.

[11] For the purpose, state block grants are defined for 1970 as the value of the matching aid for those towns for which the marginal matching rate was zero. Before 1970, *SBG* is block grants provided under the previous "foundation" plan. Although the Massachusetts rules imply that the receipt of a block grant instead of a matching grant now reflects very high or very low spending by a local community, no attempt was made to develop a simultaneous model. The insensitivity of the estimate to 1) the use of the mixed 1965–70 sample, 2) the restriction of the sample to towns with matching grants, and 3) the use of instrumental variable estimation, indicates that this is not likely to be a source of significant bias.

[12] The federal grants variable includes some state grants for special programs.

[13] The data for *RES* are based on an official 1958 survey and may for some towns be inaccurate for 1965 and 1970.

[14] The remaining small towns were omitted because data were not available for all the variables or, for a few towns, because of their relation to a regional school district.

[15] Income and wealth variables are deflated by the Consumer Price Index for Boston with 1970 equal to 1. (over)

TABLE 1—LOCAL DEMAND FOR EDUCATION

Equation	*W*	*P*	*INC*	*SBG*	*FG*	*RES*	*PRIV*	*PUP*	*GROW*	E_{-1}	R^2	Sample and Method
1.1	0.283 (0.038)	−1.000 (0.185)	0.475 (0.068)	0.066 (0.013)	0.136 (0.027)	−0.118 (0.043)	−1.112 (0.392)	0.208 (0.140)	−0.336 (0.076)		0.643	1970 Ordinary Least Squares (*OLS*)
1.2	0.256 (0.079) [0.276]	−0.940 (0.240) [−1.012]	0.445 (0.103) [0.479]	0.063 (0.015) [0.068]	0.128 (0.035) [0.138]	−0.117 (0.043) [−0.126]	−1.181 (0.428) [−1.271]	0.189 (0.146) [−0.203]	−0.325 (0.080) [−0.349]	0.071 (0.183)	0.652	1970 Instrumental Variable (*IV*)
1.3	0.211 (0.037)	−1.246 (0.232)	0.639 (0.071)	0.063 (0.017)	0.206 (0.020)	−0.055 (0.048)	−0.140 (0.440)	0.072 (0.150)	−0.528 (0.075)		0.699	1965, 1970 Pooled *OLS*
1.4	0.202 (0.047)	N.A.	0.366 (0.106)	−0.067 (0.044)	0.064 (0.030)	0.085 (0.064)	1.162 (0.630)	0.140 (0.196)	−0.265 (0.103)		0.494	1965 *OLS*
1.5	0.100 (0.035) [0.288]	N.A.	0.152 (0.008) [0.438]	0.006 (0.033) [0.017]	0.029 (0.022) [0.084]	0.044 (0.047) [0.127]	−0.118 (0.477) [−0.340]	0.197 (0.142) [0.568]	−0.232 (0.075) [−0.669]	0.653 (0.070)	0.737	1965 *IV*
1.6	N.A.	−1.016 (0.135)	0.474 (0.067)	0.067 (0.010)	0.136 (0.027)	−0.119 (0.043)	−1.127 (0.374)	0.213 (0.134)	−0.333 (0.073)		0.576	1970 *OLS* Dependent Variable Adjusted for Wealth
1.7	N.A.	−1.038 (0.093)	0.437 (0.093)	N.A.	0.167 (0.028)	−0.106 (0.035)	0.716 (0.326)	0.232 (0.117)	−0.359 (0.071)		0.787	1970; Towns with $P<1$; *OLS* Dependent Variable Adjusted for Wealth
1.8	0.361 (0.114)	−1.599 (0.746)	0.465 (0.085)	0.127 (0.042)	0.124 (0.034)	−0.145 (0.063)	−1.479 (1.235)	0.429 (0.217)	−0.285 (0.096)		0.510	1970; *IV* with *P*, *SBG*, and *PRIV* endogenous

Note: The dependent variable in all equations is real educational expenditure per pupil; in equations 1.6 and 1.7, the dependent variable is adjusted for the effect of wealth. All variables are in logarithms. See text for definitions and descriptions of samples and independent variables. Standard errors are shown in parentheses. When a lagged dependent variable is included, the long-run elasticities are shown in square brackets; constant terms are not shown; N.A. = not applicable; see text for explanation

Table 1 presents alternative estimates of the basic expenditure relation, equation (2) of Section I. Despite important differences in the methods of estimation, the estimated price and wealth elasticities of these equations are quite stable. More specifically, the price elasticity estimates are between −0.940 and −1.599 while the wealth elasticity estimates are between 0.202 and 0.361.[16]

Since the estimates are quite stable and most of the parameters are only of incidental interest in the current study, it is useful to examine first the basic estimates in equation 1.1 and then to comment on the different estimating methods used for the other equations in the table. The estimates of equation 1.1 are based on a single cross-section sample of all the towns for 1970. The elasticity with respect to wealth is very low (0.28) but the price elasticity is quite high (−1.00). The income elasticity is also quite low (0.48), casting substantial doubt on the common assertion that "... as a statistical matter, the poorer districts are financially unable to raise their taxes high enough to match the edu-

---

Educational expenditure variables are adjusted by an education price index that combines the wholesale price index and an index of New England urban teachers' salaries with common weights based on the input ratio in Massachusetts.

[16] These estimates all use the price variable implied by the legislative formula. Before 1970, the legislature did not appropriate sufficient funds to provide the matching grants implied by the legislation. If towns assumed that this incomplete funding would continue, they would act as if a different price would prevail. The calculations of this section were repeated with the price (when less than one) defined as the ratio of local reimbursible expenditure to the sum of local reimbursible expenditure plus the *actual* state matching aid. These estimates are quite similar but with slightly higher price elasticities; an appendix reporting these results is available from the author.

cational offerings of the wealthier districts" (*Serrano vs. Priest*, p. 514).[17]

The elasticity with respect to block grants (*SBG*) is very low and less than the coefficient of federal grants *FG*.[18] The negative elasticity with respect to the proportion of the tax base that is residential (*RES*) confirms that voters do regard nonresidential taxes to be less costly than residential taxes. Although measurement error[19] may have biased this coefficient toward zero, it seems safe to infer that the true elasticity is actually less than the price elasticity; this implies that voters assume that a significant part of the nonresidential tax is shifted to residents in the long run. Although a larger number of pupils attending private schools (*PRIV*) appears to depress educational spending, the coefficient is unstable and the sign varies among the equations. An increase in the ratio of pupils to population (*PUP*) appears to increase per pupil spending but the coefficient is small and insignificant. Finally, the rate of growth of the number of pupils (*GROW*) causes a reduction in per pupil spending; the coefficient (−0.34) implies that an increase of 10 percent in the five-year rate of growth would depress expenditure per pupil by less than 4 percent.

The second equation tests for a distributed lag response of local spending by adding the 1965 value of the logarithm of educational expenditures (denoted E_{-5}) to the equation. To avoid a positive bias in the coefficient of this variable,[20] the equation was estimated by an instrumental variable procedure with the lagged dependent variable treated as "endogenous." The estimates show that the lag is very short and insignificant. None of the other coefficients is appreciably different from equation 1.1.

Before the current method of state aid was introduced in 1966, there were no matching grants. In 1965, all towns faced the same price. With a pooled sample of 210 observations for 1965 and 1970, the variation in the price variable reflects the legislative change and the increases in the price of educational services[21] as well as the differences in matching rates among towns in 1970. The parameter estimates presented in equation 1.3 confirm the previous conclusions. The price elasticity is high ($\hat{\beta}_p = -1.25$) and the wealth and income elasticities are relatively low.

The 1966 legislative change provides a further method of testing whether the estimated wealth and price elasticities are biased because of the relation between price and wealth implied by the Massachusetts aid formula. Since all towns faced the same price in 1965, the estimated effect of wealth for that year would not be "contaminated" by a price effect. Equation 1.4

---

[17] Although the quoted assertion obviously refers to California, similar statements have been made about the nation as a whole. In Massachusetts, the sample town with the lowest per pupil wealth in 1970 ($10,848) spent $760 per pupil. The town with the highest per pupil wealth ($85,630) spent $1130. If the low wealth town spent as much per pupil as the high wealth town, the additional $370 per pupil would only have consumed an additional 4.2 percent of family income even if there were no additional state aid. This stands in contrast to the conclusion in the *Serrano vs. Priest* decision: "The poor district cannot freely choose to tax itself into an excellence which its tax roles cannot provide. Far from being necessary to promote local fiscal choice, the present financing system actually deprives the less wealthy districts of that option" (p. 524).

[18] The block grant elasticity actually implies a marginal propensity to consume these funds of approximately 0.6. This high value is misleading because in 1970 most block grants were paid to towns that passed the limit of matching aid; the block grants are endogenous and the coefficient is biased upwards. The estimates in the 1965 equations show that an unbiased estimate is not significantly different from zero and that the endogeneity of this variable does not affect the coefficient of any of the other variables.

[19] See fn. 13.

[20] A positive bias would occur if the disturbance term were correlated with the lagged dependent variable. In the current context, this could reflect a common town-specific component in each year's disturbance. For a full statement of such a model, see Pietro Balestra and Marc Nerlove.

[21] See fn. 11.

presents the estimated coefficients for 1965.[22] The estimated wealth elasticity (0.20) is very similar to the value in 1970.[23] Although the lagged dependent variable ($ln\ E_{1960}$) introduced in equation 1.5 implies a significant lag in the response, the long-run elasticities (shown in square brackets) are very similar to the estimates of equation 1.4. In particular, the long-run wealth elasticity is estimated to be 0.29. These estimated wealth elasticities can be treated as "extraneous" estimators of the 1970 wealth effect and used to derive new estimates of the 1970 elasticities with respect to the price and other variables. Equation 1.6 uses the long-run wealth elasticity of equation 1.5; the dependent variable is $ln\ E_{1970} - 0.288\ ln\ W_{1970}$. Since the extraneous estimator is quite similar to the unconstrained estimate of equation 1.1, this substitution has little effect on the price and other elasticities.

As a final check on the estimated price and wealth elasticities, the equation was reestimated for only those towns which effectively received matching grants in 1970; for all of the observations used to estimate equation 1.7, the price variable was less than one. Since for these towns the price is related by an almost exact formula to local wealth, an extraneous estimator of the wealth elasticity is required; the estimate of 0.202 from equation 1.4 was used. The estimated price elasticity (−1.038) and the other coefficients are quite similar to the values obtained for the full 1970 sample.

In all of these equations, the price, state block grants, and use of private schools were treated as exogenous variables. When these variables are treated as endogenous in equation 1.8, the results are not substantially changed. The price elasticity is increased and the wealth elasticity rises somewhat less. The larger standard errors suggest that these differences are not statistically meaningful.

The price elasticity will be used in the next section to assess the effects of different grants on wealth neutrality. The other coefficients of the demand equation are not relevant as such but only through their contribution to the adjusted wealth elasticity. Recall from Section I that the adjusted wealth elasticity (β_w) is defined as follows: If the basic demand equation is

$$(6) \quad ln\ E_i = \beta_0 + \beta_1\, ln\ W_i + \beta_p\, ln\ P_i + \sum_{j=3}^{k} \beta_j\, ln\ X_{ji} + \epsilon_i$$

the adjusted wealth elasticity is

$$(7) \quad \beta_w = \beta_1 + \sum_{j=3}^{k} \beta_j \gamma_{x_j w}$$

where $\gamma_{x_j w}$ is the elasticity of the variable X_j with respect to W. Although β_w could be constructed by estimating individual $\gamma_{x_j w}$ and combining them with β_j as presented in Table 1, there is a simpler and more direct method that yields exactly the same estimate of β_w. In the ordinary least squares estimate of the equation:

$$(8) \quad ln\ E_i - \beta_p\, ln\ p_i = b_0 + b_1\, ln\ W_i$$

the value of b_1 is an estimate of β_w. It follows from Theil's formula for specification bias that this estimate of β_w is exactly the same as would be obtained by direct substitution in equation (7).

The adjusted wealth elasticity has been calculated in this way for three of the values of $\hat{\beta}_p$ presented in Table 1. The basic estimate of $\hat{\beta}_p = -1.000$ based on the 1970 cross-section (equation 1.1) implies an adjusted wealth elasticity of $\hat{\beta}_w=0.370$. The pooled sample of 1965 and 1970 observations (equation 1.3) implied $\hat{\beta}_p=-1.246$, the corresponding adjusted wealth elasticity is $\hat{\beta}_w=0.414$. Finally, with the value of $\hat{\beta}_p=-1.038$ obtained in equation 1.7, the adjusted wealth elasticity is 0.377.[24] Over the relevant range, therefore, the estimated price elasticity is approximately three to five times as large as the corresponding adjusted wealth elasticity. The implications of these estimates are the subject of the next section.

---

[22] There is no coefficient for the price variable because it was a constant in 1965.

[23] The coefficient of the block grant variable is now insignificant. This reflects a correction of the bias that arose from the endogeneity of the block grant in 1970; see fn. 18. The elasticities with respect to income, federal grants, pupils per capita, and the growth rate of pupils are very similar to the 1970 values. The two remaining variables, the residential portion of the tax base and the relative number of private school pupils, are insignificant.

IV. Analysis of Alternative Matching Grant Plans

This section uses the parameter estimates to assess the effect of alternative matching grants on the degree of wealth neutrality and then discusses the tradeoff between wealth neutrality, efficiency loss, and income redistribution. It is useful to begin by restating that wealth neutrality is measured as the elasticity of expenditure per pupil with respect to wealth (local tax base per pupil). It was shown in equation (4) that the value of wealth neutrality α_1 is related to the price and adjusted wealth elasticities of demand (β_p and β_w) by

$$\alpha_1 = \beta_w + \beta_p \gamma_{pw} \tag{9}$$

where γ_{pw} is the elasticity of price with respect to wealth that is implied by the matching grant plan.

With no wealth neutralizing matching aid (i.e., $\gamma_{pw}=0$), $\alpha_1=\beta_w$: the actual elasticity of educational expenditure with respect to wealth would simply be the adjusted wealth elasticity. The estimates of β_w reported in Section III therefore indicate that in the absence of state policy the wealth neutrality measure would be between $\alpha_1=0.370$ and $\alpha_1=0.414$. The high ratio of the price elasticity to the adjusted wealth elasticity implies that matching grant plans with even relatively low elasticities of price with respect to wealth (γ_{pw}) would substantially increase wealth neutrality (reduce α_1).[25] In 1970, the Massachusetts value of γ_{pw} was only 0.18 but this was sufficient to reduce α_1 to only 0.19.

Equation (9) shows that complete wealth neutrality ($\alpha_1=0$) is achieved when $\gamma_{pw}=-\beta_w/\beta_p$. The estimated elasticities therefore imply that complete wealth neutrality would occur with γ_{pw} somewhere between 0.33 and 0.37. It is clear that if complete wealth neutrality is preferred or is mandated by the court, it could be achieved by a relatively small increase in γ_{pw}.

The estimated elasticities also show that the basic percentage equalization form of district power equalizing would have the peculiar effect of causing educational spending in towns with low wealth to be very much higher than in towns with high wealth. Since the basic district power equalizing formula implies $\gamma_{pw}=1$, the estimates of β_p and β_w imply a value of α_1 between -0.630 and -0.832. Since the average wealth in the ten wealthiest towns is approximately four times the average wealth in the ten poorest towns, these values of α_1 indicate very substantial spending differences. While these implied

---

[24] The R^2 values associated with these equations provides an overestimate of how much of the variation in expenditure per pupil is due to wealth differences net of the effect of matching grants; the R^2 values are between 0.293 and 0.305. The simple r^2 between W and EPP in 1970 was 0.359. It is clear that the evidence does not support the common assertion to the effect that "The level of school expenditure in a district is largely determined by the wealth of the local tax base" (Wise (1968), p. 121). For a detailed analysis of the relation between school spending and individual income, see the author.

[25] This is contrary to the type of conclusion reached in Benson et al., André Daniere, and B. Levin et al. However, these studies did not use estimated demand elasticities and generally relied upon an ad hoc rule of behavior to describe the response of local governments to a change in state aid. Although there is a price effect in Stern's analysis, the effect is essentially imposed by the form of the function and not estimated from the sample.

α_1 values might exaggerate the actual outcome,[26] the evidence clearly implies that power equalizing grants would not neutralize wealth but could create a stronger (but negative) relation between wealth and spending than currently exists.[27]

A. *Incomplete Wealth Neutrality*

If the courts do not require complete wealth neutrality, the state legislatures will have the option of selecting the desired degree of wealth neutrality by balancing this against the other effects of variations in the matching rate. The rest of this section discusses the primary factors that should be considered in this choice of a price formula.

The current analysis has implicitly ignored differences in tastes for education. If wealth and the taste for education are correlated, a more appropriate definition of complete wealth neutrality might be the elimination of wealth effects per se but not the association between wealth and education that is due to correlated taste differences. Since the relation between wealth and the unobservable taste cannot be estimated empirically, the adjustment for correlated taste differences could not be made precise. A positive correlation between wealth and the taste for education would however imply that complete wealth neutrality was achieved with $\alpha_1 > 0$; for a negative correlation, $\alpha_1 < 0$. Whether the definition of complete wealth neutrality should be modified in this way is clearly a matter of ethical preference and legal judgment.

Of greater importance is the effect of intergovernmental grants on allocative efficiency. Since matching grants distort the price of education faced by local towns, they result in an excess burden. The magnitude of the efficiency loss due to any particular price distortion depends on the parameters of the ***individual*** demand curves,[28] the method of local budget determination, and the opportunities for residential relocation.[29] In addition, the usual aggregate of individual consumer surplus losses would require modification if households undervalue (or overvalue)[30] education or if there are significant externalities. Despite these complexities, it seems clear that the excess burden would generally increase as wealth neutrality is increased, i.e., as γ_{pw} rises and enlarges the average price distortion.[31]

---

[26] The value of β_p might not remain constant over a much wider range of prices than was observed in the sample. Moreover, the short-run response of price *increases* might be less than the long-run response even if price decreases have the same short-run and long-run responses. The outcome would therefore depend also on the average rate of matching, i.e., on the parameter k of equation (5). Moreover, if the price were very high for wealthy towns, a substantial switch to private schools might occur.

[27] These estimated effects of *DPE* would of course be different if, instead of the basic percentage equalization form of *DPE*, the price formula were $p_i = f(t_i) \cdot W_i$; see fn. 7. The current analysis has shown that if $f(t)$ is a constant, i.e., $f'(t_i) = 0$, educational spending will decline with wealth. This implies that t_i will vary inversely with W_i. If the $f(t_i)$ is designed to "reward effort", i.e., if $f'(t_i) < 0$, the inverse relation between E_i and W_i would be made even stronger.

[28] Note that the ***individuals'*** price elasticities can be very much less than the estimated town elasticity because individuals can take the level of school spending into account in selecting their residential location.

[29] The method of budget determination by voting will generally not yield a Pareto optimal level of spending for the community in the absence of a distortionary matching grant. This point is developed by Robin Barlow and extended in comments by Yoram Barzel, Theodore Bergstrom, and Noel Edelson. It is surprising that neither Barlow nor any of the discussants of his paper consider the issue of residential location. The effect of individual choice of community is to reduce the welfare loss associated with uniform expenditure for each town. If individuals with different wealth levels choose to reside in the same town, differences in taste for education must partly offset the effect of wealth differences. Because differences among communities in their tax bases and tax rates are capitalized and reflected in the prices of residential housing, the process of "shopping for public services" by residential location is not sufficient to yield a Pareto optimal level of local spending.

[30] Although it has long been argued that parents may, out of ignorance or selfishness, spend too little on their children's education, recent research strongly suggests that within the observed range of per pupil expenditure there is no relation between school spending and educational gain (see J. S. Coleman et al., Frederick Mosteller and Daniel Moynihan). Ignorance and good intentions may cause overspending rather than underspending.

Although wealth neutralizing matching grants generally redistribute funds toward low wealth towns,[32] a careful analysis of the final incidence of the grants may show that the effect on the real distribution of income is both unintended and undesirable. The net funds received by a town will be reflected in the value of local property.[33] The primary gains would therefore accrue to landlords and homeowners in low income communities.

The issue of "horizontal equity" (i.e., the principle that individuals who are alike should be treated in the same way) has received prominent attention in discussions of educational finance. It is commonly suggested that the current system is unfair because two families with the same wealth will pay different tax rates and therefore different total tax dollars for the same education if they live in towns with different wealth per pupil. This argument is essentially illusory since differences in the tax cost of a given level of education will be largely offset by differences in housing prices between the communities. If one town offers the opportunity to obtain education at a lower cost, its houses will sell at higher prices.

Finally, the choice among matching grant plans will imply different total costs to the state government. Since these educational deficits will affect state taxes and other expenditure programs, the size of the deficit will have additional distributional and allocative effects. The size of the deficit can be controlled by changing the average price without affecting the wealth neutrality, i.e., by altering k without changing γ_{pw}. It is clear, however, that with only two "instruments" (k and γ_{pw}) it is not possible to achieve any desired combination of wealth neutrality, average expenditure, and budget deficit. More generally, if complete wealth neutrality is not judicially imposed, the state legislatures must balance wealth neutrality, allocative efficiency, and distributional impact in designing a system of educational finance.

V. Conclusion

Notions of categorical equity have become increasingly common in public discussion and judicial opinion. To incorporate such preferences in public expenditure analysis requires operational measures of categorical equity and explicit attention to balancing categorical equity, allocative efficiency, and distributional impact.

The current paper has analyzed an important example of categorical equity: wealth neutrality in local public education. For this purpose, wealth neutrality was measured by the elasticity of local school spending with respect to local per pupil wealth. The analysis showed that a state can achieve any desired degree of wealth neutrality by using matching grants that cause the net price of educational services to vary with wealth in the appropriate way.

---

[31] The excess burden would therefore also depend on the value of k in the price formula. If k were chosen to make the *maximum* price be 1, the excess burden with any $\gamma_{pw}>0$ would be greater than if k were chosen to make the *average* price be 1. By centering the prices around 1, the average absolute price distortion is less for any γ_{pw}; as a rough approximation, the excess burden is proportional to the square of the absolute price distortion. For some estimates of the welfare loss of distortionary finance, see Eric Toder.

[32] This is obviously so if the grant is completely wealth neutralizing. With partial wealth neutrality, the distribution depends on balancing the effects of higher spending by wealthy towns and higher matching rates for less wealthy towns. It is easily shown that the elasticity of state matching payments with respect to local wealth (η_{sw}) is given by $\eta_{sw}=\alpha_1[E(1-P)/S]+\gamma_{pw}[EP/S]$ where E is the expenditure per pupil and S is state aid per pupil. For sufficiently low γ_{pw}, η_{sw} may be positive. Obviously the net redistribution depends on how any state deficits are financed.

[33] Any reduction in local taxes that results from the matching aid would be fully capitalized. Each additional dollar of educational spending will be worth less than a dollar to the local residents. This value would also be reflected in the price of individual homes or, through its effect on the rents that people would pay, on the value of rental property.

The specific parameter estimates of the educational demand function imply that even complete wealth neutrality can be achieved in Massachusetts with matching grants that involve a relatively low elasticity of price with respect to wealth (between 0.33 and 0.37). The estimated elasticities therefore indicate that the widely advocated district power equalizing form of percentage equalization grant, which has a price-wealth elasticity of 1, would not be wealth neutralizing but is more likely to result instead in an *inverse* relation between local wealth and local educational spending.

There is much that can be done to verify and extend this analysis. The current calculations should be applied to other states to see whether achieving different degrees of wealth neutrality would require more substantial price distortion than in Massachusetts. The empirical analysis should also be extended by using a more complete model in which the price and quality of educational services and the value of local property are treated as endogenous. A better theoretical understanding of the process of residential location, budget determination and property valuation is required as a basis for assessing the allocative efficiency and distributional impact of alternative systems of educational finance.

Finally, the methods discussed in this paper are obviously applicable to any public service in which intergovernmental aid is intended to have a wealth neutralizing effect. The same approach can also be used in the design of matching grants to individuals. Thus, the basis of educational wealth neutrality could be shifted from the school district to the individual.[34] More generally, national health insurance, rent subsidies, and public scholarships for higher education all involve subsidies that intentionally distort the prices that individuals pay in order to achieve greater wealth neutrality. Such programs would benefit from explicit analysis of the way that alternative price systems would affect wealth neutrality as well as allocative efficiency and income distribution.

REFERENCES

K. J. Arrow, "A Utilitarian Approach to the Concept of Equality in Public Expenditures," *Quart. J. Econ.*, Aug. 1971, *58*, 409–15.

R. W. Bahl and R. J. Saunders, "Factors Associated with Variations in State and Local Government Spending," *J. Finance*, Sept. 1966, *21*, 523–34.

P. Balestra and M. Nerlove, "Pooling Cross-Section and Time-Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas," *Econometrica*, July 1966, *34*, 585–612.

R. Barlow, "Efficiency Aspects of Local School Finance," *J. Polit. Econ.*, Sept. 1970, *78*, 1028–39.

Y. Barzel, "Private Schools and Public School Finance," *J. Polit. Econ.*, Jan. 1973, *81*, 174–86.

C. S. Benson et al., *Final Report to the Senate Select Committee on School District Finance*, Vol. 1, California, June 1972.

T. Bergstrom, "A Note on Efficient Taxation," *J. Polit. Econ.*, Jan. 1973, *81*, 187–91.

G. A. Bishop, "Stimulative versus Substitutive Effects of State School Aid in New England," *Nat. Tax J.*, June 1974, *17*, 133–43.

J. M. Buchanan and C. J. Goetz, "Efficiency Limits of Fiscal Mobility: An Assessment

---

[34] In contrast to Milton Friedman's original proposal for educational vouchers of fixed value, the state could provide each household with a voucher that specifies a price at which the household could purchase educational services. With appropriate estimates of the individual wealth and price elasticities, the pricing system could be designed to achieve any desired degree of wealth neutrality. Coons et al.'s suggestion of "individual power equalization" raises the same problem as district power equalization; unless the individuals' price and adjusted wealth elasticities are equal, individual power equalizing grants will not be wealth neutralizing.

of the Tiebout Model," *J. Public Econ.*, Apr. 1972, *1*, 25–44.

J. S. Coleman et al., *Equality of Educational Opportunity*, Washington 1966.

J. E. Coons, W. H. Clune, and S. D. Sugarman, *Private Wealth and Public Education*, Cambridge 1970.

A. Daniere, *Cost-Benefit Analysis of General Purpose State School-Aid Formulas in Massachusetts*, Report to the Massachusetts Advisory Council on Education, 1969.

N. M. Edelson, "Efficiency Aspects of Local School Finance: Comments and Extensions," *J. Polit. Econ.*, Jan. 1973, *81*, 158–73.

S. Fabricant, *The Trend of Government Activity in the United States since 1900*, New York 1952.

M. Feldstein, "Income Differences and Individual Choice in Public Education," mimeo., 1974.

M. Friedman, *Capitalism and Freedom*, Chicago 1962.

E. Gramlich, "Alternative Federal Policies for Stimulating State and Local Expenditures: A Comparison of Their Effects," *Nat. Tax J.*, June 1968, *21*, 119–29.

———, "Effect of Federal Grants on State-Local Expenditures: A Review of the Econometric Literature," *Proc. 62nd Annual Conference on Taxation, Nat. Tax Assn.*, 1969.

J. M. Henderson, "Local Government Expenditures: A Social Welfare Analysis," *Rev. Econ. Statist.*, May 1968, *50*, 156–63.

R. Inman, "Three Essays on Fiscal Federalism," unpublished doctoral dissertation, Harvard Univ. 1970.

———, "Towards an Econometric Model of Local Budgeting," *Proc. 64th Annual Conference on Taxation, Nat. Tax Assn.*, 1971.

B. Levin et al., *Paying for Public Schools: Issues of School Finance in California*, Washington 1972.

F. Mosteller and D. P. Moynihan, *On Equality of Educational Opportunity*, New York 1972.

R. A. Musgrave, *The Theory of Public Finance*, New York 1959.

W. E. Oates, *Fiscal Federalism*, New York 1972.

J. W. Osman, "The Dual Impact of Federal Aid on State and Local Government Expenditures," *Nat. Tax J.*, Dec. 1966, *19*, 362–72.

E. Renshaw, "A Note on the Expenditure Effect of State Aid to Education," *J. Polit. Econ.*, Apr. 1960, *68*, 170–74.

S. Sacks and R. Harris, "The Determinants of State and Local Government Expenditures and Intergovernmental Flows of Funds," *Nat. Tax. J.*, Mar. 1964, *17*, 75–85.

D. Stern, "Effects of Alternative State Aid Formulas on the Distribution of Public School Expenditures in Massachusetts," *Rev. Econ. Statist.*, Feb. 1973, *55*, 91–97.

H. Theil, *Economic Forecasts and Policy*, 2d ed., Amsterdam 1961.

C. M. Tiebout, "A Pure Theory of Local Expenditures," *J. Polit. Econ.*, Oct. 1956, *64*, 416–24.

J. Tinbergen, *Contributions to Economic Analysis on the Theory of Economic Policy*, Amsterdam 1963.

E. Toder, "Equalizing Educational Expenditures: A Welfare Analysis of Some Plans for Reform of Spending on Public Education," unpublished paper, 1973.

A. E. Wise, *Rich Schools, Poor Schools: The Promise of Equal Educational Opportunity*, Chicago 1968.

Advisory Commission on Intergovernmental Relations, "State Financing of Public Elementary and Secondary Schools," in *ACIR State Legislative Program—New Proposals for 1970*, Washington 1969.

New York State Commission on the Quality, Cost, and Financing of Elementary and Secondary Education, *The Fleischmann Report on the Quality, Cost and Financing of Elementary and Secondary Education in New York State*, Vol. I, New York 1973.

President's Commission on School Finance, *Schools, People and Money, The Need for Educational Reform*, final rept., Washington 1972.

Rodriguez vs. San Antonio, Federal Suppl., U.S. District Court of Western Texas, Dec. 23, 1971.

——— vs. ———, 41 U.S. Law Week 4407, March 20, 1973.

Serrano vs. Priest, L.A. 29820, Superior Court No. 938254, cited in *Harv. Educ. Rev.*, Nov. 1971, *41*, 503.

Household Utility Maximization and the Working Wife

By WENDY LEE GRAMM*

There have been many studies on the labor supply of married women.[1] Typically, they find that the presence of young children in the household is an important factor affecting women's labor force behavior. Yet in many of these studies, the presence of children is not explicitly incorporated into the underlying theoretical framework. In this paper, I develop a model which incorporates children into the decision-making process. It is assumed that the household derives utility from commodities which are produced using time and market inputs.[2] In this model, the presence of a child affects production, and this influence varies with its age. Disaggregated data on over 400 households are used to test hypotheses about the influence of children on the woman's labor force behavior. The data, collected in a survey of married women teachers, are uniquely suited for this study—partly because there is much information not readily available in other bodies of household data, and partly because the sample itself has many desirable characteristics (for the purposes of testing the hypotheses implied by this model).

I. A Utility-Maximizing Model of the Household

In this model, the household maximizes its utility which depends on the amount of home-produced goods H consumed. Home goods are produced using inputs of time of the husband T_1, time of the wife T_2 and/or market inputs X. The household faces a budget constraint which states that expenditures on market goods cannot exceed income (which equals property income rA and the labor incomes of the husband and wife). The household also faces a time constraint which states that the time spent per period at market work plus the time spent per period producing home goods must exhaust the length of the period (which is arbitrarily set at one). The problem can be stated as follows:

$$(1) \qquad \text{Maximize } U(H)$$

subject to:

$$H = g(X, T_1, T_2, a_K, \tau) \quad K = 0, \ldots, L$$

$$rA + w_1(1 - T_1) + w_2(1 - T_2) - p_X X = 0$$

$$(1 - T_2) \geq 0$$

where a_K is the number of children between the ages K and $K+1$, L is the age of the child when it leaves the household, τ is the age of the household (the number of years since marriage), r is the interest rate, A is household assets, w_1 and w_2 are wages of the husband and wife, respectively, and p_X is the price of market inputs. The first relationship is the production function for home goods, the second summarizes the budget and time constraints, and the third states that the wife may choose to spend all her time at home.[3]

---

* Assistant professor of economics, Texas A&M University. I would like to thank Thomas Saving, George Borts, and Arleen Leibowitz for helpful comments, especially on the theoretical section of this paper. The empirical sections are drawn from my dissertation. I am grateful for the assistance I received in this from Frank Brechling, George Delehanty, Clyde Porterfield, Arthur Treadway, and Richard Westin. The research was supported by grants from the Manpower Administration of the U.S. Department of Labor under the Manpower Development and Training Act of 1962, as amended, and from the Texas Research Council.

[1] See, for example, Nedra Belloc, William Bowen and T. Aldrich Finegan, Glen Cain, Malcolm Cohen et al., Thomas Mahoney, Lucy Mallan, Jacob Mincer (1962a), and Richard Rosett.

[2] See Gary Becker and Reuben Gronau for a development of this model.

In this model, the presence of children affects the household decision-making process by affecting the productivity of market inputs and nonmarket time in the production of home goods. Accordingly, the marginal products of inputs may differ for households with different numbers and age distributions of children. For example, compare two households identical in all respects except that one has no children and one has a single child age one year. The difference between the two households' marginal products of inputs in the production of home goods represents the effect that a one-year old child has on the household (relative to a household with no children). The effect of the aging of a child can then be represented by comparing the marginal products for a family with one child age two years with that same family with one child age one year. No specific assumptions are made here about the relative sizes of the marginal products of inputs for utility maximizing households with children of different ages. However, one would expect the marginal products of time and certain market inputs to be greater for households with children vis-à-vis childless households. One might also expect that for households alike in all respects other than the ages of the children, the marginal products of the husband's and wife's times are largest for households with very young children while the marginal product of market inputs is largest for households with older children.[4]

The age of the household, defined as the number of years since marriage, affects the productivity of the husband and wife in producing home goods. Economists often assume that specific on-the-job training raises the marginal product of labor in that job. Also, education makes the individual a more efficient producer and consumer. For similar reasons it is assumed here that with age the husband and wife acquire skills and experience which causes the marginal products of their time in the production of home goods to increase.[5]

The following are the necessary conditions for the maximization of utility.[6]

$$(2) \quad U_H g_X - \lambda p_X = 0$$

$$(3) \quad U_H g_{T_1} - \lambda w_1 = 0$$

$$(4) \quad U_H g_{T_2} - \lambda w_2 - \theta = 0$$

$$(5) \quad rA + w_1(1-T_1) + w_2(1-T_2) - p_X X = 0$$

$$(6) \quad \theta \geq 0; \qquad = 0 \quad \text{when} \quad (1-T_2) > 0$$

where U_H is the marginal utility of H and where g_X, g_{T_1}, and g_{T_2} are the marginal products of X, T_1, and T_2, respectively. The interpretation of the necessary conditions is straightforward. Conditions (2)–(4) state that in the interior (when $\theta=0$ and $1-T_2>0$) the ratio of marginal utilities of T_1, T_2, and X must equal the ratio of prices. Conditions (2)–(4) also reduce to the condition that the marginal products in producing home goods must equal the ratio of prices. Since the multiplier λ is the marginal utility of income, these conditions imply that for an interior solution the marginal utility produced by an hour of market work must equal the marginal utility produced with an hour of home time. Equation (5) restates the budget constraint, while (6) is the boundary condition which states that when the wife does not work in the market, the multiplier θ is greater than zero. When this is so, (4) states that

---

[3] Since the purpose of this paper is to derive the wife's labor supply behavior and not the husband's labor supply, it is assumed that the husband always spends some time at market labor.

[4] One might expect this to be so because time is essential to the attainment of utility from some commodities (like "childness" or "TV watching"). So the husband's and wife's times are less substitutable for market purchased time and goods and have higher marginal products. As the children age and are more able to care for themselves (really as the children's ability to produce childness for the household increases) the marginal products of the adults' time in producing and consuming this commodity decline. See Arleen Leibowitz (1972) and the author (1974) for other discussions of this issue.

[5] See Leibowitz (1974) and Robert Michael for the effect of formal education on household decisions. Formal education tends to increase the quantity of labor supplied because it is market oriented. In other words, formal education tends to raise the productivity of labor in the market more than the productivity at home. The kind of education obtained with age is like job-specific training and increases efficiency in the production of nonmarket commodities. For example, homeowners may become expert handymen as they get older. See Mincer (1962b) for a discussion of job-specific training.

[6] The necessary conditions for the problem as it is set up in equation (1) contains another multiplier λ_2 associated with the first constraint, the production function. This multiplier has already been solved out of the system of necessary conditions given here (the multiplier λ_2 is equal to the marginal utility of home goods).

$$U_H g_{T_2} > \lambda w_2 \tag{7}$$

or when the marginal utility of an extra hour of home time exceeds the marginal utility of an extra hour of market work, the wife will not work in the market. When the wife is spending some time at market labor, then $\theta=0$, and these conditions say that the marginal utility of an extra hour of home work time just equals the marginal utility of an extra hour's pay.

The demand for the wife's home time can be derived from the necessary conditions (2)–(6) and is a function of wages, price, interest rate, household assets, household age, number of children of different ages, and parameters of the utility and production functions. The comparative statics for the interior solution are also derived by totally differentiating the necessary conditions and using Cramer's Rule. Some of the more important relations are discussed below:

(a) The sign of the income effect $(\partial T_2/\partial A)$ may be positive or negative, depending on the signs and relative sizes of the second partials of the utility and production functions. The degree to which market inputs are substitutes for time in the production of commodities is important in determining whether or not the wife desires more home time when assets increase.

(b) The effect of an increase in w_1 or w_2 on T_2 depends on the sign and relative sizes of substitution and income effects. When the husband's wage increases, it causes his nonmarket time to be more expensive. In producing home commodities, the time of the wife or market inputs may be used in place of the husband's time. If time of the wife is substitutable for husband's time and if market goods are not good substitutes for time, then the substitution effect (of an increase in the husband's wage on the wife's nonmarket time) will be positive. If the income effect is positive or smaller than the substitution effect $\partial T_2/\partial w_1$ will be positive.[7] The effect of an increase in the wife's own wage on her nonmarket time depends on relative sizes and signs of the income and substitution effects (but the own substitution effect is negative due to second-order necessary conditions).

(c) The effect of children on the nonmarket time of the wife is the primary concern of this paper. Of particular interest is the effect of a change in the number of children age K in the household given by equation (8) where U_H is the marginal

$$\frac{\partial T_2}{\partial a_K} = [-U_H g_{Xa_K} D_{XT_2} - U_H g_{T_1a_K} D_{T_1T_2} - U_H g_{T_2a_K} D_{T_2T_2}] \div D \tag{8}$$

utility of home goods; g_{Xa_K} is the change in the marginal product of input X due to a change in a_K; and D_{XT_2}/D is a substitution effect between the wife's time and market inputs in the production of home goods.[8] The coefficients $g_{T_1a_K}$, $g_{T_2a_K}$, $D_{T_1T_2}$, and $D_{T_2T_2}$ are similarly defined. It is expected that a household with a child age K will have larger marginal products of all inputs than households with no children (so $g_{T_1a_K}$, $g_{T_2a_K}$, and g_{Xa_K} are greater than zero). If the weighted own substitution effect $(-U_H g_{T_2a_K} D_{T_2T_2}/D)$ dominates (this may be due to a large value for $g_{T_2a_K}$) or if the wife's time is substitutable for other inputs, then $\partial T_2/\partial a_K$ will be positive and the wife will spend more time at home.

---

[7] The income effect of an increase in wage on the nonmarket time of the wife may be positive when there are young children in the home, especially if one of the home-produced commodities is childness, a utility-providing output which takes inputs of time and children to produce. If households don't like children, the income effect may be negative since then the wife may go to work and hire a babysitter.

The effect of the aging of a child can be derived from equation (8). If a child ages one year, then the effect on the nonmarket time of the wife would be equal to the effect of a one unit increase in a_K plus the effect of a one unit decrease in a_{K-1} or $(\partial T_2/\partial a_K - \partial T_2/\partial a_{K-1})$. We can thus derive how the labor supply behavior of the wife changes as the child ages.

(d) The effect of the age of the household on the nonmarket time of the wife is similar to the effect of children as described above. It is made up of two weighted substitution effects.

$$(9)\quad \frac{\partial T_2}{\partial \tau} = \frac{-U_H g_{T_1\tau} D_{T_1T_2} - U_H g_{T_2\tau} D_{T_2T_2}}{D}$$

I am assuming that $g_{X\tau}=0$. The substitution effects are weighted by the change in the marginal products of time due to age (times the marginal utility of home goods). One would expect $\partial T_2/\partial \tau$ to be positive implying that women of older families would tend to spend less time at market labor, *ceteris paribus*.

Elasticities of demand can be derived from the necessary conditions. Of particular interest is the wife's labor supply elasticity $\epsilon_{h_2w_2}$ which is derived from the elasticity of demand for her nonmarket time with respect to her wage $\epsilon_{T_2w_2}$. The supply elasticity is:

$$(10)\quad \epsilon_{h_2w_2} = -\frac{T_2}{1-T_2}\left[\frac{w_2(1-T_2)}{Y}\epsilon_{T_2Y} + \frac{w_2T_2}{Y}\sigma_{T_2T_2}\right]$$

where Y is full income, ϵ_{T_2Y} is the income elasticity of demand for T_2, and $\sigma_{T_2T_2}$ is the Allen partial elasticity of substitution. Equation (10) implies that the sign of the wife's labor supply elasticity depends on the relative sizes of the income and substitution effects. The size of the coefficient of elasticity depends on the amount of time the wife spends at market labor $(1-T_2)$, the size of the income elasticity of the wife's nonmarket time ϵ_{T_2Y} and the size of the elasticity of substitution in production $\sigma_{T_2T_2}$. Furthermore, if the demand for the nonmarket time of the wife (T_2) is larger for families with children, then the labor supply elasticity will be larger for families with children than for families without children, *ceteris paribus*.

In summary, the wife's labor supply decision (one minus the nonmarket time of the wife) is a function of the husband's and wife's wages, the ages of the children, household age, household assets, prices and the interest rate, and the parameters of the utility and production functions as given below:

$$(11)\quad h_2 = 1 - T_2 = h_2(w_1, w_2, A, a_K, \tau, p_X, r, \text{parameters of the utility and production functions})$$

Some of the expected signs of the partial derivatives of the labor supply equations are[9]

---

[8] D is the determinant of the matrix of coefficients of the totally differentiated system of necessary conditions and D_{ij} represents the cofactor of the ith row and jth column of this matrix where rows and columns are indexed X, T_1, T_2, λ.

[9] Relations such as $\partial h_2/\partial p_x$ are not discussed here since the labor supply equation was estimated using cross-section data and variables like prices are assumed to be the same for everyone in the sample.

$$\frac{\partial h_2}{\partial w_1} < 0 \tag{12}$$

(if the substitution and income effects of w_1 on h_2 are negative or if a positive income effect is smaller than a negative substitution effect)

$$\frac{\partial h_2}{\partial w_2} ?$$

(also depends on the size of the positive own substitution effect of an increase in w_2 on h_2 relative to the income effect which may be positive or negative)

$$\frac{\partial h_2}{\partial A} ?$$

$$\frac{\partial h_2}{\partial r} < 0$$

$$\frac{\partial h_2}{\partial a_K} < 0$$

(for households with young children relative to households with no children).

In the following section, a linear version of the labor supply equation is estimated using data particularly suited for this analysis.

II. Data and Estimation Procedure

The labor supply equation (11) was estimated using disaggregated cross-section data. The sample we used for this study is composed of 414 female teachers, substitute teachers, and women qualified to teach but not presently working. The information was collected in a survey conducted in 1970.[10] The women are married, living with their husbands, and most reside and teach in public primary schools in the northern suburbs of Chicago.[11] Our sample includes women of varying ages, with and without children, and with varying work loads. In this sample there are 249 women who worked full-time during the survey year, 115 worked part-time (substitute teaching), and 50 chose not to work at all during the year.

To estimate the labor supply of the wife, financial, demographic, and labor supply information was collected for each family. Since the sample is from a relatively homogeneous population in a small geographic area, it was assumed that all individuals face the same interest rate and prices. Because information on parameters of the utility and production functions were not available, these variables were omitted from our estimation.[12] Thus the only explanatory variables are the wage of the husband, the wage of the wife, ages of the children, household assets, and age of the household.

The measure of the amount of market work supplied by the wife (h_2) that is used in the estimation is the proportion of the full-time load that she worked in 1969. The variable h_2 is thus continuous and can assume any value from zero to one. There are observations where h_2 lies between zero and one since substitute and former teachers were also surveyed. Information was obtained about the number of days per week that a substitute taught on the average in 1969 and this was compared with the amount a regular teacher worked to obtain the proper measure of h_2 for a substitute teacher. The measure used for the wife's and husband's wages is the amount they could earn (in thousands of dollars per year) if they worked a full load every day. For substitutes this is the amount she could earn if she substituted all day every day.[13] Information on the value of present assets and the age of the household was given in the questionnaire. Understandably, many women found it difficult to estimate the value of household assets so we expect that these figures are somewhat inaccurate.

---

[10] The survey of working women was conducted by going to the schools and hiring a teacher to help distribute the questionnaires, explain the purpose of the project, and collect completed questionnaires (anonymous and returned to us in sealed envelopes). Substitute and nonworking teachers were reached via mail survey (names were obtained from the schools' administrations). See the author (1971) for a more detailed discussion of the survey.

[11] This sample has several desirable qualities. First of all, it is relatively homogeneous with respect to such variables as job conditions, taste for market work, and educational requirements for the job. Also, the labor supply decision (the desired variable) of married women teachers was then probably much closer to actual employment (the observed variable) than for married women in other occupations. The woman in this sample is much like the woman of our theoretical model: she knows the wage she can earn by working (detailed salary schedules are readily available and actual salaries of the teachers are published in local newspapers); she can get a job if she decides she wants to work; the cost of job search is negligible; she knows the conditions under which she will be working; and she has more freedom to choose the number of hours she works because she can become a substitute teacher. A teacher in this area can register to substitute in several school districts if she wishes and is not penalized for not working every time she is called.

[12] As long as the omitted variables are not systematically related to the included variables then the least squares estimates of the coefficients will be unbiased.

In order to separate out the effect of each child on the labor supply behavior of the wife, we allow a_K to be equal to zero or one only. For families with more than one child, there will be more sets of age variables, one for each child. These are really dummy variables, defined as follows:[14]

$$a_K^i = \begin{cases} 1 & \text{if the } i\text{th child is } K \text{ years old} \\ 0 & \text{otherwise} \end{cases}$$

A child is K years old if it falls into a three-year age bracket. For example, a_{11}^1 denotes the first child who is 10, 11, or 12 years old, a_5^2 denotes the second child who is 4, 5, or 6 years old. There are seven age dummies to represent the ages of the first and second child (1–3 years, 4−6 years, 7−9 years, 10−12 years, 13–15 years, 16–18 years, and greater than 18 years old). Since no family in our sample had a third child older than 18 years old, there are only six age dummies for the third child. We have five dummies to represent the age of the fourth child and four for the fifth child.

Using ordinary least squares, the work of the wife (the proportion of the full-time load she worked) is regressed on her wage, her husband's wage, household assets, age of the household, and the children's age dummies. The general form of the equation estimated is:

$$\text{(13)} \quad h = c_0 + c_1 w_1 + c_2 w_2 + c_3 A + c_4 \tau + b_2^1 a_2^1 + b_5^1 a_5^1 \ldots + b_2^2 a_2^2 + b_5^2 a_5^2 + \ldots + u$$

The regression coefficient of the age dummy b_K^i measures the effect that the ith child (age K) has on the wife's supply of labor relative to the omitted group, *ceteris paribus*. When the whole sample is used in the estimation, the omitted group is the group of families with no children. For example, b_2^1 is a measure of the difference between the work loads of a woman with one child age 2 years and a woman with no children, holding all else constant. The estimates of b_K^1 can be plotted against K to see the effect of the aging of a child on the labor supply of the wife.

III. Work Patterns of Married Women Teachers

The estimation was carried out two ways. First the labor supply equation (13) was estimated using the whole sample. The variable h_2 was then estimated three more times using subsamples of one-child, two-child and three-child families.[15] Subsamples were used to see how labor supply profiles were changed by the presence of other children in the family. Finally, additional equations were estimated to obtain more information on the effects of children on h_2 as compared to the effects of the aging of the household (or the wife).[16]

---

[13] For substitutes, this figure is usually much lower than the wage she could earn as a full-time teacher. This can be expected to influence our results in the following manner: If a sample contains both full-time and part-time teachers, the coefficient of w_2 would be overstated. Likewise, the measured own-wage supply elasticity would overstate the true elasticity. This problem as well as an attempt to measure the extent of this bias is discussed more fully in Section III.

[14] See Arthur Goldberger, pp. 218–27, for a discussion of the use of dummy variables in regressions.

Table 1 lists the regression results for the four samples. In Figure 1 the estimated age dummy coefficients are plotted for the first child only.

Overall, the variables used are quite successful in explaining the variation in the work loads of the teachers. The coefficient for the wage of the husband has the predicted sign and is statistically significant. The coefficient for the age of the household is negative and statistically significant. The estimated coefficient for the household's present assets is negative but is not statistically significant. Since it was difficult for many women to estimate the value of the household's assets and since value of household assets can be expected to be correlated with wages and the household's age, we do not reject the hypothesis that assets affect the labor supply decision of the wife on the basis of these results. More accurate estimates of household assets than we have here are necessary. Our theory did not give us any predictions about the sign of the wife's own wage coefficient. However the estimated coefficient is positive and significant. This result may reflect the choice set that a woman in this sample faces rather than the response to the change in w_2. A woman in this sample can choose to work full-time, part-time, or not at all. If her choice is not to work or to work a full-time load, her hours are determined. If she chooses to be a substitute teacher, however, she is able to vary her work load as she wishes. The wage she earns as a substitute even if she works every day is considerably lower than the wage she could earn as a full-time teacher (even as a full-time teacher at beginning wages, see fn. 13). Substitute teachers may thus be paying a premium for the privilege of working less than the full-time schedule. So the direction of causality may not be that substitutes work fewer hours because the wage is low. Rather the wage is low because they chose to substitute. In order to test this hypothesis, the regression was run using the subsample of substitute teachers. The estimated coefficient of the wife's wage was small and not statistically significant.[17]

One of the more significant results is that the estimated coefficients of the age dummies for each child in most cases increases as the age of the child increases. This implies that for this group of women, the marginal product of women's time in producing home goods is largest for households with young children.

Other results are summarized as follows. The whole-sample results indicate that the presence of the first child seems to be very important in explaining the labor supply

---

[15] The omitted variable for one-child families is a_2^1; and a_2^1 and a_2^2 are omitted for two-child families. The omitted class for three-child families is made up of families whose first two children are in the 5-year old age bracket and whose third child is in the 2-year old age bracket. This means that the coefficients for the whole-sample age dummies are not directly comparable with the coefficients of the subsample age dummies. For example, in the whole-sample regression, the effect of having a 5-year old child instead of no children is $-.398$. The effect of having a 5-year old child instead of a 2-year old first child is equal to the difference in the coefficients $[-.398-(-.525)]=+.127$. This figure can be compared to the results of the regression using the subsample of one-child families where the effect of having a 5-year old child instead of a 2-year old child (the omitted class) is 0.87.

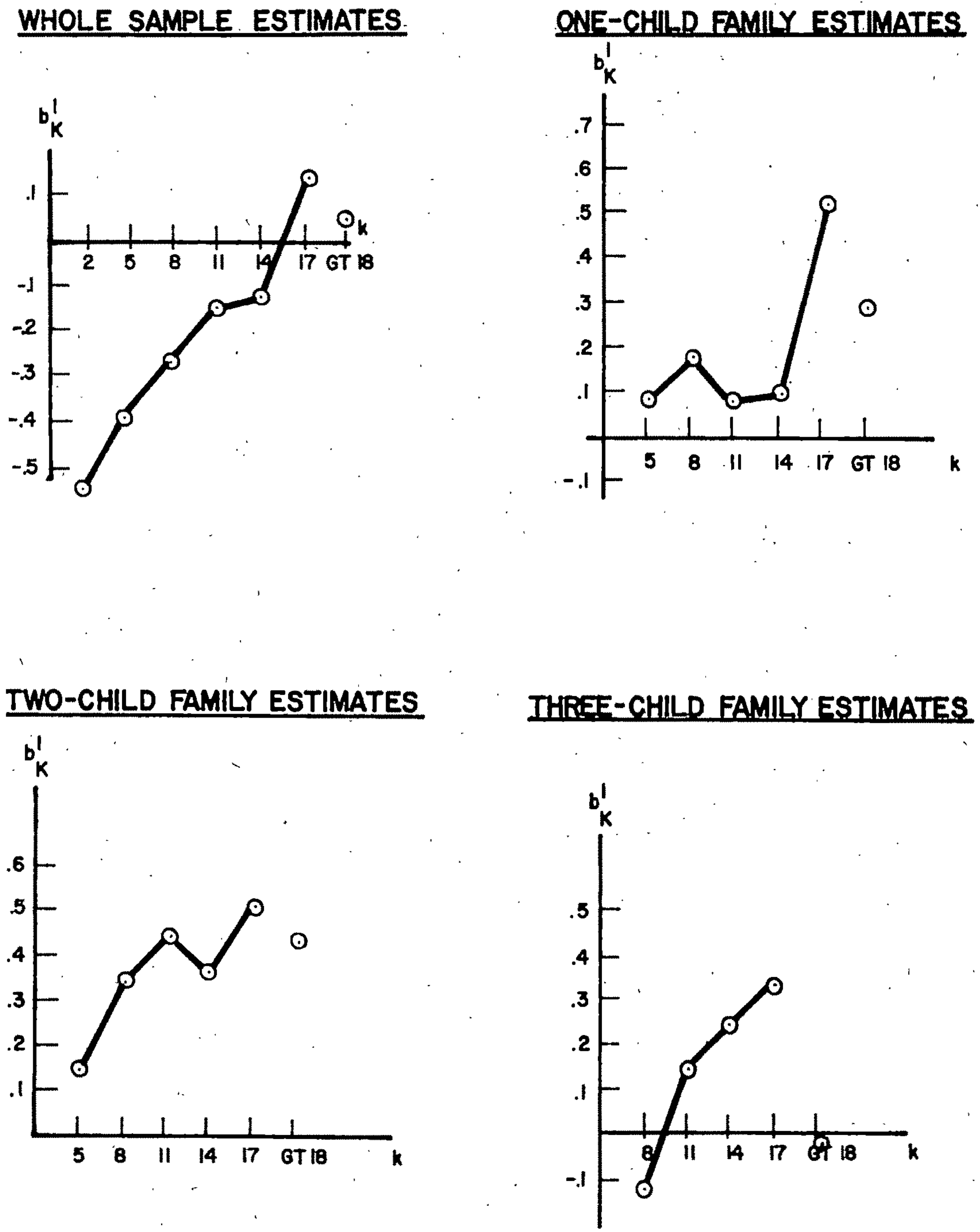
[16] In this sample, the simple correlation coefficient between the age of the household and the age of the wife is .930 and substituting the age of the wife for the age of the household in our equations did not materially affect our results.

[17] There is an additional problem in interpreting this coefficient because the variation in wages for substitute and nonworking teachers is small for this sample. All substitute teachers in a school district make the same wage.

TABLE 1—ESTIMATED REGRESSION COEFFICIENTS: WHOLE SAMPLE, ONE-CHILD FAMILY, TWO-CHILD FAMILY, AND THREE-CHILD FAMILY RESULTS

Variables	All Families	One-Child Families	Two-Child Families	Three-Child Families
Wage of Husband	−.011[c]	−.005	−.013[c]	−.008
Wage of Wife	.079[c]	.097[c]	.080[c]	.059[b]
Household Assets	−.0003	−.001	.000003	−.0001
Household Age	−.006[c]	.007	.004	−.033[b]
First Child:				
age 2	−.525[c]			
5	−.398[c]	.087	.142	
8	−.284[c]	.189	.332	−.101
11	−.152[b]	.085	.436[b]	.143
14	−.139[b]	.098	.357	.243
17	.144[a]	.561[b]	.503[b]	.338
*GT*18[d]	.048	.281	.434	−.032
Second Child:				
age 2	−.074			
5	.036		.020	
8	.062		−.017	.212
11	.035		.034	.042
14	.013		.075	.172
17	−.063		.067	.226
*GT*18	.003		−.103	.864[a]
Third Child:				
age 2	−.138			
5	−.255[b]			−.036
8	−.046			.076
11	.027			.318
14	.086			.442
17	−.214			−.007
Fourth Child:				
age 2	−.063			
5	−.317[c]			
8	−.166			
11	−.106			
14	.123			
Fifth Child:				
age 2	−.217			
5	.083			
8	.380[b]			
11	−.189			
Constant	.423	−.309	−.230	.345
R^2	.706	.666	.709	.756
F	27.691	8.785	11.112	5.535
Sample Size	414	55	90	54

[a] Regression coefficient statistically significant at .90 level of confidence.
[b] Regression coefficient statistically significant at .95 level of confidence.
[c] Regression coefficient statistically significant at .99 level of confidence.
[d] *GT*18 denotes greater than 18 years.


WHOLE SAMPLE ESTIMATES

ONE-CHILD FAMILY ESTIMATES

TWO-CHILD FAMILY ESTIMATES

THREE-CHILD FAMILY ESTIMATES

FIGURE 1. PLOTS OF ESTIMATED COEFFICIENTS OF AGE DUMMIES FOR FIRST CHILD

decision of the wife. All coefficients except b^1_{17} and b^1_{GT18} are significantly different from the omitted class at at least the .95 level of confidence. The effect of the aging of this child is also marked.[18] The second child does not seem to have a large effect on the wife's labor supply decision. The coefficients of the age dummies for the second child are all close to zero and the T-values are all small. This may be due to the fact that the first child has picked up some of the effect of the second child. Or it may be due to economies of scale to housework or a learning-by-doing effect with respect to child care.

Another interesting result is found by comparing the estimated coefficients of the different subsamples. The mother of two or three children returns to work more quickly than a mother of one child. This is not surprising since mothers of large families may respond to the increased need for income by returning to work while her children are still young. This rapid increase slows down (there is a "flattening out") when the first child is older perhaps because of the presence of the younger children.

Elasticities of supply were evaluated at the mean for the subsamples. As predicted by the theory, the own-wage elasticity of supply is smallest for the subsample of families with no children. (The elasticities were .610 for families with no children, 1.473 for one-child families, 1.025 for two-child families and .728 for three-child families.)[19] Our theory does not provide an explanation for the negative relationship between the number of children in the household and the size of the coefficient of elasticity.

In order to explore the relationship between the ages of the children and the age of the household, several more equations were estimated using the whole sample. First, the age of the household was expressed as dummies, similar to the dummies representing the children's ages. As a first step, the explanatory powers of different groups of variables were compared, with the following results. When the equation contained only wages and assets as independent variables, R^2 was .520. When household-age dummies were added to the equation, R^2 was .550. However, when children-age dummies were added to the equation in place of the household-age dummies, R^2 was .700. Thus the coefficient of partial determination (partial R^2) or relative net contribution of the household-age dummies was .062 with an F-ratio of 2.600.[20] This is to be compared with the partial R^2 for the children-age dummies of .375 with an F-ratio of 7.655. This implies that the addition of children-age dummies added much more towards explaining the variation in h_t than the addition of household-age dummies.

Furthermore, inspection of the results of the original full sample estimates reported in Table 1 reveals that the estimated coefficient of the household's age (τ, not dummies) though small in magnitude is nonetheless significantly different from zero at the .99 level of confidence. This indicates that the age of the household is a factor affecting the labor supply decision of the married female in this sample and this factor is indeed distinct from the effect of the various ages of her children.

---

[18] T-tests were performed on the estimated coefficients of the age dummies to see whether or not coefficients representing the ith child are significantly different from each other. For the first child, the coefficients of dummies representing adjacent ages are usually not significantly different from each other. For example, the effect of the two-year old first child is not significantly different from that of the five-year old child. Dummies for nonadjacent ages, however, are generally significantly different from each other at the .95 level. This lends further support to the view that the effect of a child on the wife's supply of labor does indeed depend on his age. The age effect is not as strong for the other children.

[19] Since all subsamples included both substitute and full-time teachers, these elasticities may be overestimated.

[20] See Goldberger and Emanuel Melichar.

REFERENCES

G. Becker, "A Theory of the Allocation of Time," *Econ. J.*, Sept. 1965, *75*, 493–517.

N. B. Belloc, "Labor Force Participation and Employment Opportunities for Women," *J. Amer. Statist. Assn.*, Sept. 1950, *45*, 400–10.

W. G. Bowen and T. A. Finegan, *The Economics of Labor Force Participation*, Princeton 1969.

G. G. Cain, *Married Women in the Labor Force: An Economic Analysis*, Chicago 1966.

M. S. Cohen, S. A. Rea, Jr., and R. I. Lerman, *A Micro Model of Labor Supply*, Bur. Labor Statist. staff pap. 4, Washington 1970.

A. S. Goldberger, *Econometric Theory*, New York 1964.

W. L. Gramm, "The Demand for the Wife's Non-Market Time," *Southern Econ. J.*, July 1974, *41*, 124–33.

———, "A Model of the Household Supply of Labor Over the Life Cycle: The Labor Supply Decision of Married School Teachers," unpublished doctoral dissertation, Northwestern Univ. 1971.

R. Gronau, "The Intrafamily Allocation of Time: The Value of the Housewife's Time," *Amer. Econ. Rev.*, Sept. 1973, *63*, 634–51.

M. Kosters, "Income and Substitution Parameters in a Family Labor Supply Model," unpublished doctoral dissertation, Univ. Chicago 1966.

A. Leibowitz, "Education and the Allocation of Women's Time," in F. T. Juster, ed., *Education, Income and Human Behavior*, New York 1975.

———, "Women's Allocation of Time to Market and Non-Market Activities: Differences by Education," unpublished doctoral dissertation, Columbia Univ. 1972.

T. Mahoney, "Factors Determining the Labor Force Participation of Married Women," *Ind. Lab. Relat. Rev.*, July 1961, *14*, 563–77.

L. B. Mallan, "Financial Patterns in Households with Working Wives," unpublished doctoral dissertation, Northwestern Univ. 1968.

E. Melichar, "Least-Squares Analysis of Economic Survey Data," in *1965 Proc. Amer. Statist. Assn., Bus. Econ. Statist. Sec.*, Washington 1966, 373–385.

R. T. Michael, "Education's Effect on Consumer Behavior," presented at the Second World Congress of the Econometric Society, Cambridge 1970.

J. Mincer, (1962a) "Labor Force Participation of Married Women: A Study of Labor Supply," in *Aspects of Labor Economics*, Universities–Nat. Bur. Econ. Res. conference series 15, Princeton 1962, 63–105.

———, (1962b) "On-the-Job Training: Costs, Returns and Some Implications," *J. Polit. Econ.*, Oct. 1962, *70*, 50–79.

R. N. Rosett, "Working Wives: An Econometric Study," in T. F. Dernburg, R. N. Rosett, and H. Watts, eds., *Studies in Household Economic Behavior*, New Haven 1958, 53–99.

A Dynamic Analysis of Taxation

By OSVALDO H. SCHENONE*

In a paper published in this *Review*, Jacob Stockfisch uses a two-sector model of a competitive economy which devotes a constant fraction of income to capital accumulation to show, *inter alia*, that "A general tax on consumer items makes it possible for a given level of money investment to buy more real physical assets. Economic growth will be encouraged and the economy will have a larger physical stock of capital with the passage of time" (p. 298).

One of the purposes of this paper is to present an alternative model which gives the result that a general tax on consumption is dynamically neutral, leaving the path of capital accumulation unaffected.

Stockfisch's result hinges on the peculiar savings behavior he assumed. It can be shown that models of capital accumulation in which savings is assumed to be equal to a certain fraction of income (such a fraction may be either constant or a function of the interest rate and/or the distribution of income) give the result that consumption tax favors capital accumulation and growth.[1] This result, however, does not hold in the intertemporal optimizing model employed in this paper.

Nonetheless, a consumption tax is sometimes proposed to foster capital accumulation and growth along the lines suggested by Stockfisch. As stated by R. J. Chelliah, "When we look upon commodity taxation as a weapon for promoting economic growth, its justification lies in the fact that it has a tendency to restrain consumption" (p. 86). Or in the words of John Due, "The tax structure must reduce private consumption. . . . Thus, the tax structure must be designed to strike the portion of income spent on consumption" (p. 33).

On the other hand, arguments have been produced showing that a consumption tax would not have any effect on the path of capital accumulation. This is essentially Nicholas Kaldor's contention about the neutrality of the expenditure tax. Accordingly, the replacement of an income tax by an expenditure tax will increase savings. Richard Goode writes, "It is usually taken for granted that this policy will promote savings, but the basis of the belief is not obvious and is worthy of attention" (p. 232).

The preceding exposition points out the necessity of a systematic analysis of the effects of taxation in the long run, which is precisely the topic of this paper. In the context of an intertemporal optimizing model, I analyze the welfare implications of tax-induced changes in the path of capital accumulation and show the dynamic neutrality of a consumption tax. The model is also used to compare the welfare costs of income taxation and investment taxation.

The results show that the welfare cost per dollar of tax revenue of income taxation is smaller than the welfare cost per dollar of tax revenue of investment taxation. It is also shown that an investment tax at a constant-over-time rate is equiva-

---

* Visiting professor of economics at the Universidad Católica de Chile. This paper is based on part of my Ph.D. dissertation. I wish to acknowledge my gratitude to the members of my thesis committee, Stanley Fischer, William Brock, and Stephen Magee for guiding me throughout this study. I would also like to thank Arnold Harberger and George Borts for their comments and suggestions. A referee's comments were also very helpful.

[1] This has been shown in the first chapter of my dissertation.

lent to a consumption subsidy at a decreasing-over-time rate.

I. An Intertemporal Optimizing Model

A. *The Long-Run Welfare Cost of Taxation*

The welfare cost of taxation arises from the fact that people are induced to make their decisions according to relative prices that do not reflect the true choices technology and factor endowments make available. This is quite applicable to several types of choices: 1) choices among several goods at a point in time; 2) choices among the same good at different points in time; and 3) choices among different goods at different points in time.

Most of the existing analyses of the welfare cost of taxation take a static point of view according to which utility- and profit-maximizing individuals are confronted *at a point in time* with an exogenously imposed wedge between marginal rates of transformation and marginal rates of substitution.[2] This is an accepted procedure for short-run analysis. It is not, however, fully satisfactory from a long-run point of view and a concern exists about the necessity of introducing dynamic considerations in the analysis of the welfare cost of taxation. As Arnold Harberger puts it: "As long as we take our problem seriously, we must face up to a possible effect of taxation on the volume of savings and investment and thus the problem of incorporating a dynamic element in what would otherwise be a static model must be faced" (p. 58).

The same concern can be found much earlier in the literature and may even be the basis of John Stuart Mill's disatisfaction with the double taxation on savings implied by the income tax. More recently, Irving Fisher and Kaldor took Mill's point to further elaborate that an income tax distorts the savings-consumption choice, and thus has an effect on the path of capital accumulation. This argument *inter alia* leads Kaldor to propose the adoption of an expenditure tax as a perfectly neutral tax regarding the savings-consumption choice; hence, the path of capital accumulation.

In agreement with this concern, we want to recognize that from a dynamic point of view, the welfare cost arises from distortions of all three types of choices mentioned above. That is, in the long run the welfare cost arises not only from an inefficient use of given resources, but also from the fact that the amount of resources available at every point in time itself is nonoptimal. It is my purpose to analyze the total welfare effects of taxation over the whole lifetime of taxes which are assumed to be enforced for a period long enough to affect the factor endowments of the economy. In so doing I will formally show the dynamic neutrality of Kaldor's expenditure tax and the dynamic nonneutrality of the income tax. The starting point of the analysis must be to consider the capital stock as an endogenous variable subject to the influence of taxation.

If we consider the capital stock as an endogenous variable, it seems very convenient to use an aggregation that separates the production of I goods from the production of other goods (consumption goods, C) and makes clear whether the changes in taxation under consideration strike I goods or C goods. The aggregation used in the standard two sector model à la Hirofumi Uzawa seems to be appropriate for this purpose. The supply side of the model has linearly homogeneous production functions for both commodities.

$$(1) \quad I(t) = F_I(K_I(t), N_I(t)) = N_I(t) f_I(k_I(t))$$

$$(2) \quad C(t) = F_c(K_c(t), N_c(t)) = N_c(t) f_c(k_c(t))$$

---

[2] One exception is a paper by David Levhari and Eytan Sheshinski. They analyze the lifetime welfare cost imposed by an interest-income tax upon a single individual whose income consists of a return (at a constant-over-time rate) on his capital.

We also assume full employment of capital K and labor N at every point in time t.

$$K_I(t) + K_c(t) = K(t) \tag{3}$$

$$N_I(t) + N_c(t) = N(t) \tag{4}$$

Assuming no depreciation, the aggregate quantity of capital existing at any point in time is given by the accumulation of I goods which have been produced in the past; namely,

$$DK(t) = I(t) \tag{5}$$

where D operator denotes time derivative.

We also assume that labor grows at a constant rate n,

$$DN(t)/N(t) = n$$

Because in a two-sector model factor allocations and output levels are fully determined by the only relative price of the model p, and the capital-labor ratio k, the procedure we will follow consists of evaluating the tax-induced changes in the pairs (p, k) that the economy attains at every point in time. This is analogous to the concern in static analysis with the effect of a tax upon market equilibrium (i.e., how equilibrium quantity and price change in response to a change in taxation). Once we know how the path is distorted, we search for the welfare implications of these changes in the path. This is analogous to the static procedure of evaluating the welfare triangles, once it is known how the instantaneous market equilibrium is affected by a change in taxation.

The procedure will be illustrated with a simple, particular case. On the demand side of the model we will postulate the simplest aggregate utility function; namely, utility equals per head consumption, $q_c(t)$ $(\equiv C(t)/N(t))$

$$U(t) = q_c(t)$$

This is the simplest case and for this reason I want to use it for my present illustrative purpose.[3]

Following Uzawa, assume the economy behaves so as to

$$\text{Maximize} \int_0^\infty U(q_c(t))e^{-\delta t}dt$$

subject to the constraints (1) to (5), where δ is the discount rate by which future consumption is weighted against present consumption. Allocating Lagrange multipliers $p_I(t)$, $p_c(t)$, $r(t)$, $w(t)$, and $q(t)$ to the constraints (1) to (5), yields the following Euler equations

$$U'(t) - p_c(t) = 0 \tag{6}$$

$$q(t) - p_I(t) = 0 \tag{7}$$

$$p_c(t)\partial F_c/\partial K_c(t) - r(t) = 0 \tag{8}$$

$$p_I(t)\partial F_I/\partial K_I(t) - r(t) = 0 \tag{9}$$

$$p_c(t)\partial F_c/\partial N_c(t) - w(t) = 0 \tag{10}$$

$$p_I(t)\partial F_I/\partial N_I(t) - w(t) = 0 \tag{11}$$

$$r(t) - (n + \delta)q(t) + Dq(t) = 0 \tag{12}$$

The transversality condition is

$$\lim_{t\to\infty} e^{-\delta t}q(t)k(t) = 0$$

By writing equations (6) to (12) as equalities, we are here assuming that all solutions to these Euler equations lie in the interior of the feasible set. This assumption rules out specialization. Clearly, $p_c(t) = U'(t)$ is the price of C goods; $p_I(t)$, which is equal to $q(t)$ by (7), is the price of I goods; $r(t)$ is the competitive rental rate of capital (from (8) and (9)); and $w(t)$ is the competitive wage rate (from (10) and (11)). To describe the optimal path we take the two dynamic relationships of the model; namely, equations (5) and (12). Now rewrite equation (5) in per capita terms

---

[3] It has been shown by the author that the results to be obtained here also hold in a more general model in which $U''<0$ (and, consequently, the locus $Dp=0$ is not a horizontal line).

$$
\begin{aligned}
&(13) \quad & Dk(\mathrm{t}) &= q_I(\mathrm{t}) - nk(\mathrm{t}) \\
&\text{where} & k(\mathrm{t}) &= K(\mathrm{t})/N(\mathrm{t}) \\
& & q_I(\mathrm{t}) &= I(\mathrm{t})/N(\mathrm{t}) \\
& & &= \frac{k_c(\mathrm{t}) - k(\mathrm{t})}{k_c(\mathrm{t}) - k_I(\mathrm{t})} f_I(k_I(\mathrm{t}))
\end{aligned}
$$

From (8), (9), (10), and (11) we know that factor intensities in the nonspecialization region are functions of the relative price only. Choosing the consumption good as the numeraire, define $p(\mathrm{t})$, the relative price of investment goods in terms of consumption goods, as $p_I(\mathrm{t})/p_c(\mathrm{t})$. Then equation (13) becomes

$$(14) \quad Dk(\mathrm{t}) = q_I(p(\mathrm{t}), k(\mathrm{t})) - nk(\mathrm{t})$$

Equations (7), (9), and (12) imply

$$
\begin{aligned}
\frac{Dp_I(\mathrm{t})}{p_I(\mathrm{t})} &= (n + \delta) - f_I'(k_I(\mathrm{t})) \\
&= (n + \delta) - f_I'(k_I(p(\mathrm{t})))
\end{aligned}
$$

Given the particular utility function, $q_c(\mathrm{t}) = U(\mathrm{t})$, $U'(\mathrm{t}) = 1$ and we can write

$$(15) \quad \frac{Dp(\mathrm{t})}{p(\mathrm{t})} = (n + \delta) - f_I'(k_I(p(\mathrm{t})))$$

Equations (14) and (15) form a system of two differential equations in $p(\mathrm{t})$ and $k(\mathrm{t})$. If I goods are relatively labor intensive, the locus of (p, k) along which (14) is at rest is upward sloping. On the other hand, (15) is independent of k in the nonspecialization region; i.e., (15) is at rest for the value of p that makes $(n+\delta) - f_I'(k_I(p)) = 0$.

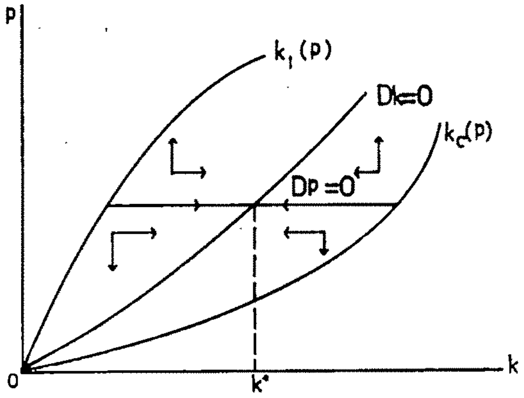


Figure 1. The Long-Run Equilibrium
(C goods are relatively capital intensive)

The rest point of the system (14) and (15) is shown in Figure 1, where k^* represents the steady-state capital-labor ratio. To keep the simplicity of this illustration, let us concentrate on the case in which I goods are relatively labor intensive. How do the time paths of k, $k(\mathrm{t})$, and of q_c, $q_c(\mathrm{t})$ look? To obtain these paths make use of the fact that p and therefore factor proportions in each sector are constant all along the path. Therefore, we can write both q_c and q_I as linear functions of k

$$q_I = \frac{f_I(k_I)}{k_I - k_c} k - \frac{f_I(k_I) k_c}{k_I - k_c}$$

$$(16) \quad q_c = -\frac{f_c(k_c)}{k_I - k_c} k + \frac{f_c(k_c) k_I}{k_I - k_c}$$

Substituting the first of the preceding two equations into (13) yields

$$Dk = \left(\frac{f_I}{k_I - k_c} - n \right) k - \frac{f_I k_c}{k_I - k_c}$$

Solving this differential equation for a given initial $k(0) = k_0$ gives

$$
\begin{aligned}
(17) \qquad k(\mathrm{t}) &= \left(k_0 - \frac{f_I k_c}{f_I - n(k_I - k_c)} \right) \\
&\quad \cdot \exp \left(\frac{f_I - n(k_I - k_c)}{k_I - k_c} \mathrm{t} \right) \\
&\quad + \frac{f_I k_c}{f_I - n(k_I - k_c)}
\end{aligned}
$$

$$(18) \quad \lim_{\mathrm{t} \to \infty} k(\mathrm{t}) = \frac{f_I k_c}{f_I - n(k_I - k_c)} = k^*$$

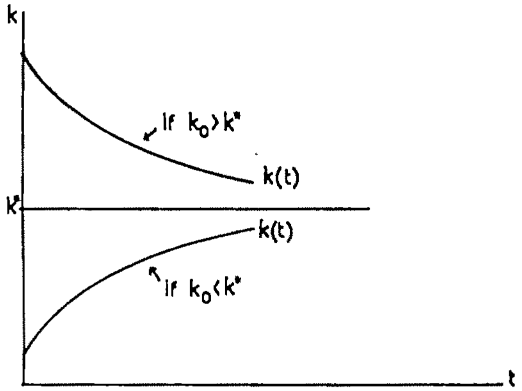


FIGURE 2. THE PATH OF CAPITAL ACCUMULATION

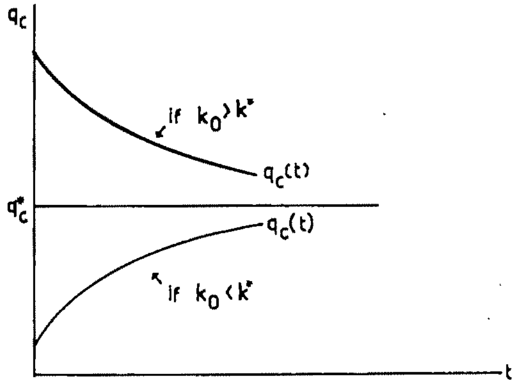


FIGURE 3. THE PATH OF CONSUMPTION

Therefore,

$$k(t) = (k_0 - k^*) \cdot \exp\left(\frac{f_I - n(k_I - k_c)}{k_I - k_c}\, t\right) + k^*$$

If $k_0 > k^*$, then k declines steadily toward k^* as t increases. If $k_0 < k^*$, then k increases steadily toward k^* as t increases. Substituting (17) into (16) gives the path of q_c, $q_c(t)$.

$$(19) \quad q_c(t) = \frac{f_c(k_I - k^*)}{k_I - k_c} - \frac{f_c(k_0 - k^*)}{k_I - k_c} \cdot \exp\left(\frac{f_I - n(k_I - k_c)}{k_I - k_c}\, t\right)$$

Equation (19) gives the optimal path of q_c, which starts at

$$(20) \quad q_c(0) = \frac{f_c(k_I - k_0)}{k_I - k_c}$$

and converges to

$$q_c^* = \frac{f_c(k_I - k^*)}{k_I - k_c}$$

These paths can be represented as in Figures 2 and 3.

Since we assume $U(t) = q_c(t)$, equation (19) is the path of utility. The discounted sum of all consumption per head along the optimal path gives the total utility of the whole program, W.

$$W = \int_0^\infty \frac{f_c(k_I - k^*)}{k_I - k_c} \exp(-\delta t)\, dt - \int_0^\infty \frac{f_c}{k_I - k_c}(k_0 - k^*) \cdot \exp t\left(\frac{f_I - n(k_I - k_c)}{k_I - k_c} - \delta\right) dt$$

Solving the above integral and making some algebraic manipulations we get

$$(21) \quad W = f_c \frac{f_I - (n + \delta)k_I + \delta k_0}{\delta(f_I - (n + \delta)(k_I - k_c))}$$

In the absence of any tax, $R = f_I' = n + \delta$; consequently, the total welfare of the program equals initial per head wealth: The present value of all future wages w plus the value of the initial capital labor ratio

$$W(\text{without tax}) = w/\delta + pk_0$$

B. *Effects of Taxation on the Time Path Followed by the Economy*[4]

In the preceding section we confronted the economy with the constraints implied by technology and factor endowments (equations (1) to (5)). In this section the situation in the presence of taxes will be described. The effect of taxes (as pointed out earlier) is to confront producers and consumers with different signals through the markets. Producers respond to a price different from the price to which the buyers respond. To make clear the effect of taxation on the signals perceived by producers and consumers, define $p_c^{\$}$ and $p_I^{\$}$ as the prices that producers receive in terms of an arbitrary numeraire (other than consumption goods, say "dollars"). Let the ratio $p_I^{\$}/p_c^{\$}$ be denoted by $\bar{p}$ (i.e., the relative price of I goods in terms of C goods as perceived by producers), and keep the notation p to denote the relative price as paid by buyers ($\bar{p}=p(1-T)$ for a tax on I and $\bar{p}=p/(1-T)$ for a tax on C). Disposable income of any individual in the model is the reward to his work $w(\bar{p})$, plus the reward to the capital owned by him $kr(\bar{p})$, plus whatever lump-sum transfer he receives from the government Z. (We follow the standard procedure in excess burden analysis of assuming that tax revenue is returned to the private sector as lump sum transfers.) When each individual decides how to spend his income, $w(\bar{p})+r(\bar{p})k+Z$, between present consumption and capital accumulation he faces a relative price p between q_I and q_c. That is, his decision may be posed as how to

---

[4] It is, of course, understood that all variables are functions of time t although the t is omitted for simplicity.

$$\text{Maximize} \int_0^\infty q_c(t)e^{-\delta t}dt$$

subject to one of the following constraints: For a tax on C at rate T, we have

$$(22) \quad \frac{p_c^{\$}}{1-T}q_c + p_I^{\$}q_I = w^{\$}(\bar{p}) + r^{\$}(\bar{p})k + Z^{\$}$$

where $p_c^{\$}$ is the price received by the producer and $p_c^{\$}/(1-T)$ is the price paid by the buyer, all expressed in dollars, and $w^{\$}(\bar{p}), r^{\$}(\bar{p})$, and $Z^{\$}$ have the same meaning as $w(\bar{p})$, $r(\bar{p})$, and Z above, except that they are now measured in dollars.

For a tax on I we get

$$(23) \quad p_c^{\$}q_c + \frac{p_I^{\$}}{1-T}q_I = w^{\$}(\bar{p}) + r^{\$}(\bar{p})k + Z^{\$}$$

For a tax on both C and I, i.e., an income tax, we obtain

$$(24) \quad \frac{p_c^{\$}}{1-T}q_c + \frac{p_I^{\$}}{1-T}q_I = w^{\$}(p) + r^{\$}(p)k + Z$$

Notice that in this last equation factor rewards are functions of p rather than $\bar{p}$. In the presence of an income tax the relative price between C and I faced by consumers equals the one faced by producers. After having made clear how the prices perceived by consumers and producers are distorted, we can now express our constraints in the more compact notation of relative prices only. From equation (22)

$$q_c + \frac{p_I^{\$}}{p_c^{\$}/1-T}q_I = \frac{w^{\$}(\bar{p})}{p_c^{\$}/1-T} + \frac{r^{\$}(\bar{p})}{p_c^{\$}/1-T}k + \frac{Z^{\$}}{p_c^{\$}/1-T}$$

Recalling that p is the relative price that a buyer faces in the market,

$$(25) \quad q_c + pq_I = (w(\bar{p}) + r(\bar{p})k + Z)(1-T)$$

Likewise, from (23)

$$(26) \quad q_c + pq_I = w(\bar{p}) + r(\bar{p})k + Z$$

And from (24)

$$(27) \quad q_c + pq_I = (w(p) + r(p)k + Z)(1-T)$$

1. *Consumption Tax*

In the presence of a tax on C goods, the problem faced by an optimizing individual is to maximize $\int_0^\infty q_c e^{-\delta t}dt$, subject to equation (25). The solution to this problem is

$$(28)\quad Dp/p = (1-T)r(\bar{p})/p - (n+\delta)$$
$$= R - (n+\delta)$$

Thus the individual will only be in equilibrium if the modified rate of time preference equals the net rate of return; i.e., the net rental he will receive per unit of capital divided by the price he paid for the capital good. The net rental he will receive is net in the sense of being determined by the net price (the price that producers receive, $\bar{p}$) and also in the sense that $(1-T)$ is taken away from it, since the individual consumes the rental. That is, when the individual considers the purchase of a machine, as opposed to immediate consumption, he deducts from the machine's rental the tax he will have to pay in order to consume these rentals.

Equation (28) defines the $Dp=0$ locus, which is a horizontal line as in Figure 1. The height of this line, however, is not the same as the height of the line in Figure 1 because in that case, p has to satisfy $r(p)/p=n+\delta$, whereas in the presence of a tax p has to satisfy $(1-T)r(\bar{p})/p=R=n+\delta$.

The equilibrium condition in our case is stated by equation (28); that is, by the equalization of the net rate of return $(1-T)r(\bar{p})/p$ with $n+\delta$. This is, of course, obviously true in the steady state, as can be seen by making equation (28) equal to zero. But since p is constant all along the optimal path, so is R. Then if $R=n+\delta$ in the steady state, R must be equal to $n+\delta$ all along the optimal path as well. The introduction of taxation affects the net rate of return without changing $n+\delta$. We must, therefore, know how the optimal path changes in the process of bringing back the net rate of return to equality with $n+\delta$. That is, we want to evaluate the horizontal shift of the $Dp=0$ locus due to a change in taxation. Using the fact that $E_{rT}=[T/(1-T)]E_{rp}$ and totally differentiating $(1-T)r(\bar{p})/p=n+\delta$, we get

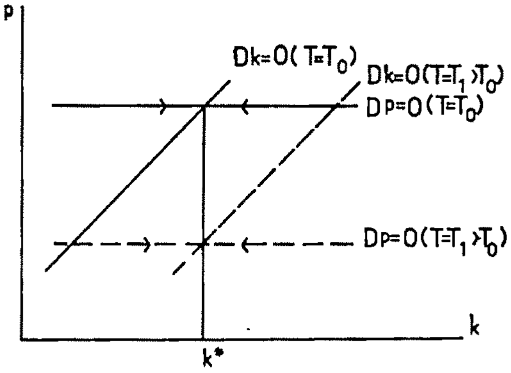


FIGURE 4. THE EFFECT OF A CONSUMPTION TAX ON THE LONG-RUN EQUILIBRIUM

$$\frac{T}{p}\frac{dp}{dT} = E_{pT} = -T/(1-T)$$

Moreover, assuming that $Dk=0$, and using the fact that $E_{q_I p}=[T/(1-T)]E_{q_I T}$ we get from $q_I(\bar{p}, k)=nk$ that

$$E_{pT} = -T/(1-T)$$

The effect of increasing a tax on C is depicted on Figure 4. Thus, an increase in the tax rate on C shifts both $Dk=0$ and $Dp=0$ schedules by the same distance, so that the steady-state capital-labor ratio k^* remains unchanged and p is reduced so as to keep $(1-T)r(\bar{p})/p=n+\delta$. The welfare implications of these shifts are discussed below.

2. *Tax on I Goods*

In the presence of a tax on I, the problem faced by an optimizing individual is to maximize $\int_0^\infty q_c e^{-\delta t}dt$, subject to equation (26). The solution to this problem is

$$(29)\quad r(\bar{p})/p = n+\delta$$

which states that the individual is in equilibrium only if the net rate of return equals the modified rate of time preference. To find out how this equilibrium condition is affected by changes in taxation, totally differentiate equation (29) to get

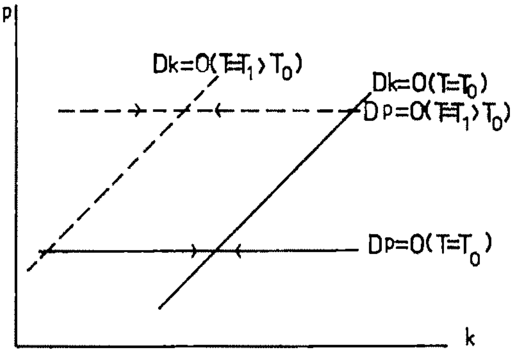


FIGURE 5. THE EFFECT OF AN INVESTMENT TAX ON THE LONG-RUN EQUILIBRIUM

$$E_{pT}\Big|_{Dp=0} = (T/1 - T)(E_{rp}/E_{rp} - 1)$$

Since $q_I(\bar{p}, k) = nk$, we get

$$E_{pT}\Big|_{Dk=0} = T/(1 - T)$$

The effect of increasing a tax on I is shown in Figure 5. Thus, an increase in the tax on I increases p and reduces $\bar{p}$, so as to keep $R = r(\bar{p})/p$ constant and equal to $n+\delta$.

How are the paths of k, $k(t)$, and of q_c, $q_c(t)$ affected by this change in taxation? From equations (18) and (20) calculate equations (30) and (31). Thus we have

$$\frac{dq_c(0)}{dT}\begin{cases} =0 & \text{for a tax on } C \\ >0 & \text{for a tax on } I \end{cases}$$

where

$$l_c(0) = \frac{N_c(0)}{N(0)} = \frac{q_c(0)}{f_c}$$

From $q_c^* = q_c(k^*, \bar{p})$ and $nk^* = q_I^* = q_I(k^*, \bar{p})$, we get

$$\text{(32)}\qquad \frac{dq_c^*}{d\bar{p}} = \frac{\partial q_c^*}{\partial k^*}\frac{dk^*}{d\bar{p}} + \frac{\partial q_c^*}{\partial \bar{p}}$$

$$= \frac{\partial q_c^*}{\partial k^*}\frac{(\partial q_c^*/\partial \bar{p})(1/p)}{(\partial q_I^*/\partial k^*) - n} + \frac{\partial q_c^*}{\partial \bar{p}}$$

$$= \frac{\partial q_c^*/\partial \bar{p}}{\partial q_I^*/\partial k^* - n}\left(\frac{r}{p} - n\right) > 0$$

That is, the steady-state level of per capita consumption moves in the same direction as the net (perceived by producers) relative price of capital $\bar{p}$. Therefore, q_c^* falls as a tax on I is increased.

It is clear from equation (30) that the steady-state capital-labor ratio k^* falls as we increase the tax rate on I, a result already observed in Figure 5. Since $\bar{p}$ falls at every time t as the tax on I is increased, q_I and therefore k will be smaller at every t. Therefore, we can show the effects of changes in taxation upon the path of k as in Figure 6. (Only the case $k_0 < k^*$ is shown. The opposite case follows *mutatis mutandi.*)

From equation (30) it can be seen that fixed proportions in both sectors will prevent any effect of taxation upon k^*. This is intuitively clear since in the presence of fixed proportions, and because of the full employment assumption, the economy will always be at a kink on the transformation curve and no changes in the composition of output are to be expected from "small" changes in taxation. With no

---

$$\text{(30)}\qquad \frac{dk^*}{dT} = \frac{\dfrac{dk_I}{dT}((k_c - k^*)f_I' + k^*n) + \dfrac{dk_c}{dT}(f_I - nk^*)}{f_I - n(k_I - k_c)} < 0$$

$$\text{(31)}\qquad \frac{dq_c(0)}{dT} = \left[\frac{dk_I}{dT}(f_c - q_c(0)) + \frac{dk_c}{dT}(l_c(0)(f_c - k_cf_c') + f_c'l_c(0)k_I)\right] \div [k_I - k_c]$$

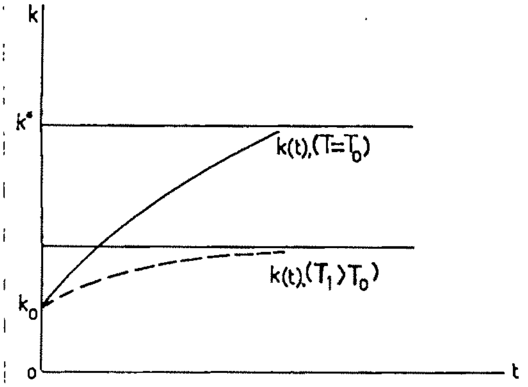


FIGURE 6. THE EFFECT OF AN INVESTMENT TAX ON THE PATH OF CAPITAL ACCUMULATION

change in the composition of output, the path of k remains unchanged.

Exactly the same comments apply to equation (31), which yields the change in per capita consumption at the time the tax rate changes, $t=0$. This result is also intuitively clear: as the tax on I is increased, the economy moves on the transformation curve towards more C being produced and consumed. This result together with equation (32) (which contains the effect of taxation upon steady-state per capita consumption q_c^*) allows the representation of the effect of taxation upon the path of q_c as in Figure 7.

Note that the path of q_c is twisted by a change in taxation. That is, an increase in the tax on I induces an increase in present consumption which implies less accumulation of capital, and results in less consumption in the future. These tax-induced modifications of the time path of consumption entail departures from the optimal path and thus a welfare cost which we discuss below.

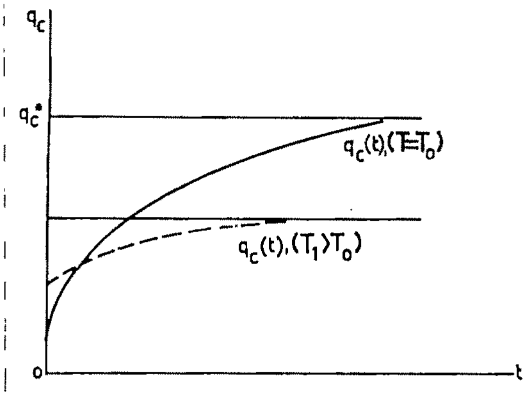


FIGURE 7. THE EFFECT OF AN INVESTMENT TAX ON THE PATH OF CONSUMPTION

3. *Income Tax*

In the presence of an income tax the problem faced by an optimizing individual is to maximize $\int_0^\infty q_c e^{-\delta t}dt$, subject to equation (27). The solution to this problem is

$$(33) \qquad (1 - T)r(p)/p = n + \delta$$

Which we totally differentiate with respect to p and T to get

$$E_{pT}\bigg|_{Dp=0} = (T/1 - T)(1/E_{rp} - 1) < 0$$

Equation (14) is not affected by income taxation, so that

$$E_{pT}\bigg|_{Dk=0} = 0$$

The final outcome is shown in Figure 8. Thus, an increase in the income tax rate reduces p so as to keep $(1-T)r(p)/p$ constant and equal to $n+\delta$. This reduction in

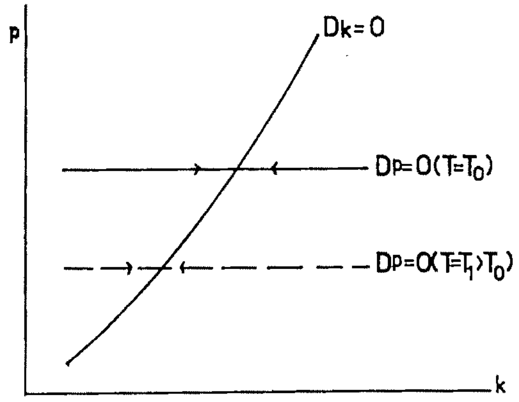


FIGURE 8. THE EFFECT OF AN INCOME TAX ON THE LONG-RUN EQUILIBRIUM

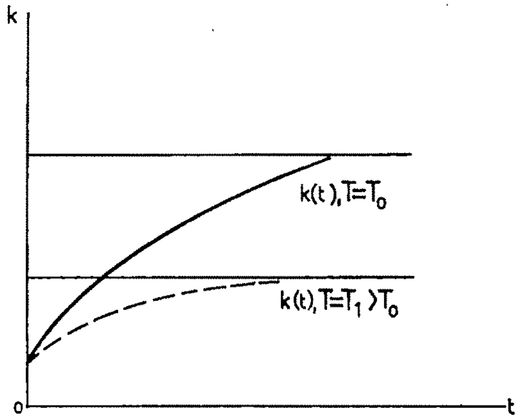


FIGURE 9. THE EFFECT OF AN INCOME TAX ON THE PATH OF CAPITAL ACCUMULATION

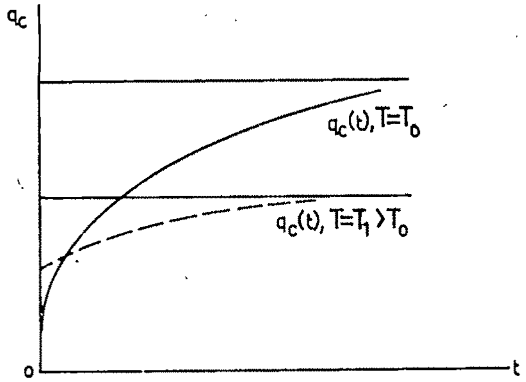


FIGURE 10. THE EFFECT OF AN INCOME TAX ON THE PATH OF CONSUMPTION

p implies (since p is now identical to $\bar{p}$) a fall in $\bar{p}$, which produces a reduction in q_I. This results in a lower k, a lower k^*, and a lower q_c^* (see equations (30) and (32)) as well as a greater $q_c(0)$ according to equation (31). Therefore, the paths of k and q_c are affected as shown above in Figures 9 and 10. The twist in the consumption path entails a welfare cost which we discuss next.

C. *Welfare Implications of Tax-Induced Changes in the Path*

1. *Consumption Tax*

Since a consumption tax will strike both present and future consumption, consumers face an unchanged relative price between present and future consumption. Therefore, their intertemporal maximizing behavior still consists of allocating their resources exactly as they would in the absence of taxation.

They cannot, however, buy future consumption as such. They can only buy investment goods in order to consume the rentals in the future. This fact brings in two effects: 1) the rental that an investment good is able to yield is itself a function of taxation (i.e., r depends on T, given p); and 2) the consumption of those rentals is taxed at rate T (unless the individual never consumes those rentals, but in such a case the purchase of the investment good becomes purposeless!). At a given p these two effects make the purchase of investment goods less attractive as T goes up. On the other hand, under the same circumstances the production of investment goods becomes more attractive ($\bar{p}$ goes up as T goes up, for a given p). There will, therefore, be an excess supply of I which will disappear as p falls. The fall in p necessary to restore equilibrium is such that $\bar{p}$ returns to its level prior to the change in taxation, so that $r(\bar{p})$ is kept unchanged. Such a fall in p leaves as the tax burden of purchasing an investment good only the tax on the consumption of the rentals (but the rentals themselves are unaffected).

Since this is also the tax burden of immediate consumption, no incentive arises to modify the allocation of any individual's resources between present and future consumption, so that present and future consumption will be demanded in the same proportion as before the change in taxation. A constant $\bar{p}$ implies that they will also be supplied in the same proportions as before the change in taxation and therefore there is neither excess demand nor excess supply at this lowered p and unchanged $\bar{p}$.

This is exactly the situation described in Figure 4. Since $\bar{p}$ is not affected by taxation the flow supply of and demand for investment goods and therefore Dk, q_c^*, and k^* are independent of taxation. Since the paths of q_c and k are undistorted, no welfare cost arises from this tax.

2. *Tax on I*

As depicted in Figure 7, an increase in the tax rate on I goods affects the intertemporal allocation of resources decreasing future and increasing present consumption. The total welfare cost of this distortion can be expressed as the change in W from equation (21).

Therefore, equation (21) becomes

$$\text{(34)}\quad W\,(\text{tax on } I) = \frac{w/\delta + pk_0 + Tpf_I/\delta}{1 + Tpf_I/f_c}$$

Now we can answer the question about the welfare implications of the tax-induced shifts of the optimal path. Take the general welfare indicator, equation (21), and differentiate with respect to the tax rate on I goods to get

$$\text{(35)}\quad \frac{1}{W}\,\frac{dW}{dT} = \frac{dk_c}{dT}\left(\frac{f_c'}{f_c} - \frac{(n+\delta)p}{f_c + Tpf_I}\right) + \frac{dk_I}{dT} \cdot \left(\frac{((n+\delta)k_c - \delta k_0)(f_I' - (n+\delta))}{(f_I - (n+\delta)k_I + \delta k_0)(f_c + Tpf_I)}\right) < 0$$

It can be seen from equation (35) that fixed proportions in both sectors will prevent the existence of any welfare cost of taxation. This is intuitively clear and follows from the fact that fixed proportions will prevent any effect of taxation upon the paths of k and q_c, as pointed out above. It can also be seen from this equation that a tax on C has no welfare cost, as already stated: One need only notice that both dk_c/dT and dk_I/dT are identically zero for a tax on C since such a tax does not change $\bar{p}$ and therefore cannot change either k_c or k_I.

3. *Income Tax*

Figure 10 shows the effect of the income tax on the path of q_c. It is analogous to the effect of a tax on I. The reason for this analogy is clear: both an income tax and a tax on I make future consumption more expensive relative to immediate consumption.

The effect of an income tax is not only analogous to the effect of a tax on I. They are also quantitatively equivalent. The reason for this is that the distortion in the path is exactly the same in both cases: as a consequence of an income tax the price received by producers falls, according to a total differentiation of (33), by $T/(1-T)(E_{rp}-1)$. As a consequence of a tax on I, the price paid by consumers rises according to total differentiation of (29), by $TE_{rp}/(1-T)(E_{rp}-1)$. Thus, the change in the price received by producers due to I tax is

$$\frac{T}{1-T}\left(\frac{E_{rp}}{E_{rp}-1} - 1\right).$$

which is equal to $T/(1-T)(E_{rp}-1)$. This is exactly the change in the price perceived by producers due to an income tax.

This numerical equivalence follows from the fact that the consumption tax is neutral. Therefore, the addition of it to a tax on I (to make up a tax on income) does not change the effect of the tax on I. They are, thus, the effects of an income tax as well.

An income tax at rate T and a tax on I at rate T have the same welfare cost. The tax revenue, however, will not be the same, an income tax has a wider base and then its revenue will be greater than the revenue of a tax on I at equal rates. Therefore, the welfare cost per dollar of tax revenue of an income tax is smaller than the welfare cost per dollar of tax revenue of an investment tax.

D. *Tax Equivalence*

It is interesting to note that the effects of taxes on the two goods in the model are not symmetric: A tax on I reduces $k(t)$ at every future t; a tax on C does not. A tax on I increases $q_c(0)$; a tax on C does not. Taxes on both C and I do not cancel each other out and thus have an effect on the economy although the relative price p is unaffected. These results may seem to conflict with the principle that only relative prices matter. There is, however, no conflict. Consumers are only concerned about consumption goods at different points in time (which are the only arguments entering their utility functions) and the principle that only relative prices matter fully applies to them. Thus, a tax on C which is kept forever leaves the relative prices of present and future consumption unchanged, implying no intertemporal allocative effect. A tax on I, on the other hand, makes the postponement of consumption (through the purchase of an investment good whose yield will be consumed) more expensive relative to immediate consumption. An income tax also makes the postponement of consumption more expensive relative to immediate consumption, even though both consumption and investment goods are taxed at the same rate (i.e., p is unaffected) because present consumption entails paying the tax only once, while future consumption entails paying a tax as the investment good is bought and again when its yield is consumed.

Since taxes on C and I are not symmetric, a tax (subsidy) on I is not equivalent to a subsidy (tax) on C: the former distorts the choice between present and future consumption while the latter does not. Thus a question arises as to what (if any) set of taxes and/or subsidies exactly duplicates the effects of another set of taxes and/or subsidies. To answer this question, it is convenient to use a very simple model. Assume a one-sector model in which the only commodity can be used for consumption or investment purposes. The economy is initially endowed with a quantity K_1 of such commodity. For simplicity assume no depreciation occurs. Keep the notation δ to represent the rate of time preference and also keep a simple utility function $U_t = C_t$, where C_t is the quantity consumed in period t.

To further simplify, assume a constant-over-time labor force which is inelastically supplied at each point in time. The economy is assumed to

$$\text{Maximize } (C_1 + C_2/(1+\delta) + C_3/(1+\delta)^2 + \ldots)$$

subject to

$$\begin{aligned} C_1 + (K_2 - K_1) &= f(K_1) \\ C_2 + (K_3 - K_2) &= f(K_2) \\ &\ldots \\ C_i + (K_{i+1} - K_i) &= f(K_i) \end{aligned}$$

where $f(K_i)$ is a neoclassical production function, namely,

$$f(K_i) = AK_i^a, \qquad A > 1,\ 0 < a < 1 \tag{36}$$

The solution to this problem is

$$f'(K_i) = \delta \tag{37}$$

which shows that the optimal stock of capital at each point in time is the one for which the marginal product of capital equals the rate of time preference.

From (36) and (37):

$$K_i = (aA/\delta)^{a/(1-a)}, \qquad i \geq 2$$

Therefore,

$$\begin{aligned} C_i &= A(aA/\delta)^{a/(1-a)}, \qquad i \geq 2 \\ C_1 &= AK_1^a - ((aA/\delta)^{1/(1-a)} - K_1) \end{aligned}$$

Assume now that a tax on investment at rate T is imposed in the first period, to be kept forever. The constraints of the problem now become

$$C_1 + (K_2 - K_1)(1 + T) = f(K_1) + Z_1$$
$$C_2 + (K_3 - K_2)(1 + T) = f(K_2) + Z_2$$
$$\dots$$
$$C_i + (K_{i+1} - K_i)(1 + T) = f(K_i) + Z_i$$

where Z_i represents the lump sum return of the tax proceeds to the private sector. The solution to the optimization problem subject to those new constraints is

$$f'(K_i) = (1 + T)\delta$$

Therefore,

$$(38) \quad K_i^T = (aA/(1 + T)\delta)^{1/(1-a)}, \quad i \geq 2$$

$$(39) \quad C_i^T = A(aA/(1 + T)\delta)^{a/(1-a)}, \quad i \geq 2$$

$$(40) \quad C_1^T = K_1^a - \left\{(aA/(1 + T)\delta)^{1/(1-a)} - K_1\right\}$$

The index T attached to the variables above denotes the optimal values of these variables in the presence of the tax.

The effect of an investment tax is to increase immediate consumption (i.e., $C_1^T > C_1$) and to reduce both the steady-state capital stock (i.e., $K_i^T < K_i$, $i \geq 2$) and future consumption (i.e., $C_i^T < C_i$, $i \geq 2$). These results confirm our earlier results.

Next assume a set of subsidies to consumption is imposed. Denote by X_t the rate of the subsidy at time t. In the presence of the subsidy the constraints of the problem become

$$\frac{C_1}{1 + X_1} + (K_2 - K_1) = f(K_1) - Z_1$$
$$\frac{C_2}{1 + X_2} + (K_3 - K_2) = f(K_2) - Z_2$$
$$\dots$$
$$\frac{C_i}{1 + X_i} + (K_{i+1} - K_i) = f(K_i) - Z_i$$

where Z_i represents a lump sum tax equal to the amount of the subsidy. The solution to the optimization problem subject to these new constraints is

$$(1 + X_{i-1})(1 + \delta) = (f'(K_i) + 1)(1 + X_i)$$

which reduces to equation (37) if $X_{i-1} = X_i$. This confirms the dynamic neutrality of a constant-over-time rate subsidy (or tax) on consumption.

Then

$$(1 + X_{i-1})\delta = (1 + X_i)f'(K_i) + X_i - X_{i-1}$$

Therefore we have optimal values for K_i, C_1, and C_i as shown in (41), (42), and (43). The index S attached to the variables denotes the optimal values of those variables in the presence of the subsidy. By comparing C_1^S, C_i^S for $1 \geq 2$ and K_i^S for $i \geq 2$ with C_1^T, C_i^T for $i \geq 2$ and K_i^T for $i \geq 2$, respectively, we find the sequence of X_i which yields the same paths of C and K as the investment tax at constant rate T.

$$(41) \quad K_i^S = \left(\frac{aA(1 + X_i)}{(1 + X_{i-1})\delta - (X_i - X_{i-1})}\right)^{1/(1-a)} \qquad i \geq 2$$

$$(42) \quad C_1^S = AK_1^a - \left(\left(\frac{aA(1 + X_2)}{(1 + X_1)\delta - (X_2 - X_1)}\right)^{1/(1-a)} - K_1\right)$$

$$(43) \quad C_i^S = A\left(\frac{aA(1 + X_i)}{(1 + X_{i-1})\delta - (X_i - X_{i-1})}\right)^{a/(1-a)} - \left(\left(\frac{aA(1 + X_{i+1})}{(1 + X_i)\delta - (X_{i+1} - X_i)}\right)^{1/(1-a)} - \left(\frac{aA(1 + X_i)}{(1 + X_{i-1})\delta - (X_i - X_{i-1})}\right)^{1/(1-a)}\right)$$

It can be seen from equations (38) and (41) that K_i^a and K_i^T are the same if and only if,

$$(1 + T)\delta = \frac{1 + X_{i-1}}{1 + X_i}\delta - \frac{X_i - X_{i-1}}{1 + X_i}$$

That is,

$$\frac{X_i - X_{i-1}}{1 + X_i} = -T\frac{\delta}{1 + \delta} \tag{44}$$

Thus a sequence of consumption subsidies at decreasing rates (i.e., $X_{i-1} > X_i$) that satisfies equation (44) has the same effect on the path of capital accumulation as an investment tax at a constant-over-time rate T. Equation (44) is also the necessary and sufficient condition for C_1^T to be equal to C_1^S: equations (40) and (42) are the same if, and only if, equation (44) is satisfied.

Equation (44) is also the necessary and sufficient condition for C_i^T $(i \geq 2)$ and C_i^S $(i \geq 2)$ to be the same: when $(X_i - X_{i-1})/(1 - X_i) = T(\delta/1 + \delta)$, the right-hand side of equation (43) becomes equal to

$$A(aA/(1 + T)\delta)^{a/(1-a)}$$

as in equation (39).

We then conclude that a sequence of consumption subsidies at a decreasing-over-time rate satisfying equation (44) yields the same paths of consumption and capital accumulation as an investment tax at a constant-over-time rate.

REFERENCES

R. M. Bird and I. Oldman, *Readings on Taxation in Developing Countries,* Baltimore 1967.

R. J. Chelliah, *Fiscal Policy in Underdeveloped Countries,* London 1969.

J. F. Due, "Requirements of a Tax Structure in a Development Economy," in R. M. Bird and I. Oldman, eds., *Readings on Taxation in Developing Countries,* Baltimore 1967.

I. Fisher, "Double Taxation of Savings," *Amer. Econ. Rev.,* Mar. 1939, *29,* 16–33.

R. Goode, "Taxation of Savings and Consumption in Underdeveloped Countries," in R. M. Bird and I. Oldman, eds., *Readings on Taxation in Developing Countries,* Baltimore 1967.

A. C. Harberger, "Taxation, Resource Allocation and Welfare," in *The Role of Direct and Indirect Taxes in the Federal Revenue System,* Princeton 1964.

N. Kaldor, *An Expenditure Tax,* London 1955.

D. Levhari and E. Sheshinski, "Life-Time Excess Burden of a Tax," *J. Polit. Econ.,* Jan./Feb. 1972, *80,* 139–48.

J. S. Mill, *Principles of Political Economy,* New York 1965.

O. H. Schenone, "A Dynamic Analysis of Taxation," unpublished doctoral dissertation, Univ. Chicago 1973.

J. A. Stockfisch, "The Capitalization and Investment Aspects of Excise Taxes Under Competition," *Amer. Econ. Rev.,* June 1954, *44,* 287–300.

H. Uzawa, "Optimal Growth in a Two Sector Model of Capital Accumulation," *Rev. Econ. Stud.,* Feb. 1964, *31,* 7–24.

A Geometrical Exposition of Input-Output Analysis

By DANIEL A. GRAHAM*

A number of important theorems have been developed regarding the qualitative properties that input-output and related economic models must possess in order to be viable. That a number of apparently different notions of viability (for example, all industries can make profits, balanced growth is possible, any set of final demands can be produced) turn out to be equivalent, is one of the more important results. These theorems, normally expressed in the vernacular of the mathematician, have enjoyed only limited recognition, however, despite their proven usefulness and the ever increasing interest in such models. The strategy of this paper will be to exploit the purely geometrical nature of one condition for viability in order to expose the equivalence of seven other important conditions. It is hoped that this relatively simple geometrical exposition will not only tie the main mathematical results together but will also bring these results within the reach of a wider audience.

The input-output model is specified by the (semipositive) matrix of input coefficients[1]

$$A = [a_{ij}] = \begin{bmatrix} a_{11} & a_{12}, \dots, a_{1n} \\ a_{21} & a_{22}, \dots, a_{2n} \\ \cdot & \cdot \qquad\quad \cdot \\ \cdot & \cdot \qquad\quad \cdot \\ a_{n1} & a_{n2}, \dots, a_{nn} \end{bmatrix}$$

where a_{ij} denotes the amount of the ith good required as an input to produce each unit of the jth good. The jth column of A corresponds, therefore, to the input requirements of the jth industry (per unit of output). The gross output levels are given by the vector

$$x = (x_1 \quad x_2, \dots, x_n)'$$

where x_i denotes the gross output of the ith good. Notice, in this context, that the vector Ax represents the input requirements associated with gross output levels equal to x and, thus, the vector

$$x - Ax = (I - A)x$$

represents the net output levels corresponding to x.

A prime concern in input-output analysis is the question of whether or not a particular set of final demands can be satisfied. In other words, in what circumstances is the following "weak solvability" condition satisfied?[2]

1) Weak Solvability: $(I-A)x \geqq c$ has a solution $x \geqq 0$ for *some* set of final demands $c > 0$.

It is also of interest to determine the conditions under which any set of final demands could be satisfied.

---

* Associate professor of economics, Duke University. Financial support from the Ford Foundation-sponsored Duke University research project on input-output analysis is gratefully acknowledged. I am indebted to an anonymous referee for helpful comments on a previous draft of this paper.

[1] The matrix A is semipositive if every row and column is a semipositive vector. A semipositive vector is a vector composed of nonnegative elements, at least one of which is strictly positive. Since semipositive matrices form the basis of the von Neumann model and other linear economic models as well as the input-output model, the methods of this paper can be easily extended to more general circumstances.

[2] We employ the conventional vector notation that $a = b$ iff $a_i = b_i$ for all i, $a \geqq b$ iff $a_i \geqq b_i$ for all i, $a \geq b$ iff $a \geqq b$ and $a \neq b$, and $a > b$ iff $a_i > b_i$ for all i.

2) Strong Solvability: $(I-A)x=c$ has a solution $x\geqq 0$ for *any* $c\geqq 0$.

A related issue involves the question of whether or not there exist prices at which all industries could operate profitably.

3) Profitability: $p[I-A]>0$ has a solution $p>0$.

A final concern involves the dynamic specification of an input-output model that outputs of one period become the inputs of the next period, i.e.,

$$Ax_t \leqq x_{t-1}$$

denote the input constraints for production in period t. The issue in this context is to determine whether or not growth is possible.

4) Growth: $Ax_t \leqq x_{t-1}$, $x_t > x_{t-1}$ has a solution x_t, $x_{t-1} \geqq 0$.

The fact that these conditions take the form of restrictions upon the matrix A is obvious. The fact that these conditions are equivalent to one another and to any of the following additional conditions is less obvious.[3]

5) Hawkins-Simon: The leading principal minors of $[I-A]$ are all positive, i.e.,

$$\det(1-a_{11})>0,$$

$$\det\begin{bmatrix} 1-a_{11} & -a_{12} \\ -a_{21} & 1-a_{22} \end{bmatrix} > 0, \ldots,$$

$$\det[I-A]>0$$

6) Dominant Diagonal: The matrix $[I-A]$ has a dominant positive diagonal, i.e., there exist numbers (weights) $d_j>0$ such that

$$d_j(1-a_{jj}) > \sum_{i\neq j} d_i a_{ij}, \qquad j=1,\ldots,n$$

7) The dominant root of A is less than one.[4]

Condition 7) gives rise to interest in the dominant root of semipositive matrices. The most important theorems in this regard may be stated as follows:[5]

If A is a semipositive (indecomposable)[6] *matrix and if λ^* and x^* denote the dominant root and the associated characteristic vector, then:*[7]

$$\begin{array}{lll} \text{(a)} & \lambda^* \geqq 0 & (\lambda^* > 0) \\ \text{(b)} & x^* \geq 0 & (x^* > 0) \\ \text{(c)} & \dfrac{\partial \lambda^*}{\partial a_{ij}} \geqq 0 & \left(\dfrac{\partial \lambda^*}{\partial a_{ij}} > 0\right) \\ & & i,j=1,\ldots,n \\ \text{(d)} & [\lambda I - A]^{-1} \geq 0 & \textit{iff } \lambda > \lambda^* \\ & ([\lambda I - A]^{-1} > 0 & \textit{iff } \lambda > \lambda^*) \end{array}$$

(e) *x^* is not necessarily the only semipositive characteristic vector. (x^* is the only semipositive characteristic vector.)*

(f) $[\lambda I - A]$ *has all leading principal minors positive iff* $\lambda > \lambda^*$.

The purpose of this note is to explain geometrically conditions 1) through 7) and theorems (a) through (f). This explanation will be illustrated in the context of a two-good model ($n=2$) where the matrix A is indecomposable. The arguments can be immediately generalized to three goods ($n=3$) and offer considerable insights into the general case, however. The modifications that are necessary when A is decomposable require no significant changes in approach and will be obvious to the reader.

---

[3] For proofs of these propositions see Hukakane Nikaido.

[4] A number λ (possibly complex) and a vector x (possibly complex) are referred to as a characteristic root and vector, respectively, of the matrix A iff $Ax=\lambda x$. The dominant root is the root of largest absolute value (the square root of the sum of the squares of the real and imaginary parts of the complex root). In the case of a semipositive matrix this root happens to be real and nonnegative and the associated characteristic vector is real.

[5] These results are referred to as the Perron-Frobenius Theorems. See Edwin Burmeister and A. Rodney Dobell, pp. 437–40 for a compact statement of these and related results.

[6] A matrix is decomposable (indecomposable) if there exists (does not exist) a partition of the set of indices $\{1, 2, 3, \ldots, n\}$ into nonempty sets I and J such that $a_{ij}=0$ $i\in I, j\in J$. If such a partition exists it means that the industries in J can operate independently of the industries in I. If no such partition exists then every industry requires either directly or indirectly the output of every other industry.

[7] Parentheses enclose those results which depend upon A being indecomposable as well as semipositive.

The following result will be indispensable in this endeavor:[8] If a, $b \neq 0$ then $a \cdot b = |a|\,|b| \cos\theta$, where θ is the included angle between the vectors a and b. Thus:

$a \cdot b > 0$ iff a and b form an acute angle $(0 \leq \theta < 90°)$

$a \cdot b = 0$ iff a and b form a right angle $(\theta = 90°)$

$a \cdot b < 0$ iff a and b form an obtuse angle $(90° < \theta \leq 180°)$

The system that will be illustrated has the input coefficients matrix

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

We will assume initially $A > 0$ and that a_{11}, $a_{22} < 1$ ($a_{12} = 0$ or $a_{21} = 0$ would correspond to A being decomposable and $a_{ii} < 1$ means that the input requirement of an industry for its own output is less than the industry's total output).

Let us begin by considering the columns of the matrix

$$[I - A] = \begin{bmatrix} 1 - a_{11} & -a_{12} \\ -a_{21} & 1 - a_{22} \end{bmatrix}$$

The first column has the sign pattern $(+-)$ and thus belongs to the fourth quadrant while the second column has the sign pattern $(-+)$ and thus belongs to the second quadrant. Beyond this observation there are two interesting possibilities: 1) the cone generated by the two column vectors contains *all* of the positive (first) quadrant, or 2) the cone generated by the two column vectors contains *none* of the positive quadrant.[9] The first possibility is illustrated in Figure 1 and the second in Figure 2 (the cone is depicted by the shaded area in both cases). In both illustrations, the first column $(I-A)_{\cdot 1}$ is labeled B_1 and the second column $(I-A)_{\cdot 2}$ is labeled B_2.

In Figure 1 the cone B_2OB_1 contains all of the positive quadrant. The only other possibilities are that B_2 lies in the extended line OB_1 and thus that the cone B_2OB_1 is simply this extended line (the case in which $[I-A]$ is singular) or that the cone B_2OB_1 contains the negative (third) quadrant (illustrated in Figure 2). There are simply no other possibilities for the two vectors: they either "bend" toward the positive quadrant or they do not bend at all or they bend toward the negative quadrant. If there were three goods, the analogous (three dimensional) cone would either contain the positive orthant (see Figure 3), be a plane or a line, or contain the negative orthant.[10] Similarly, with n goods the cone either contains the positive orthant, is a linear subspace,[11] or contains the negative orthant. Notice that in all cases the cone either contains *all* of the positive orthant or *none* of the positive orthant.

It can now be shown that conditions 1) through 7) are equivalent by arguing that each is equivalent to:

8) The cone generated by the columns of $(I-A)$ contains the positive orthant, i.e.,

---

[8] See any text in linear algebra, for example, G. Hadley: The scalar product of two vectors, a and b, is denoted by $a \cdot b = \sum_{i=1}^{n} a_i b_i$. The Euclidean norm of a vector b is denoted by $|b| = (\sum_{i=1}^{n} b_i^2)^{1/2}$.

[9] The cone generated by a set of vectors $B_1 \ldots B_m \in R^n$ is defined to be the set $\{y \in R^n \mid y = \sum_{i=1}^{m} \alpha_i B_i, \alpha_i \geqq 0\ i = 1, \ldots m\}$.

[10] The positive orthant is the set $\{y \in R^n \mid y > 0\}$. The negative orthant is the set $\{y \in R^n \mid y < 0\}$.

[11] A linear subspace is a set S that has the property that if $x_1, x_2 \in S$ then $\alpha_1 x_1 + \alpha_2 x_2 \in S$ for all $\alpha_1, \alpha_2 \in R$ (all real numbers α_1 and α_2).

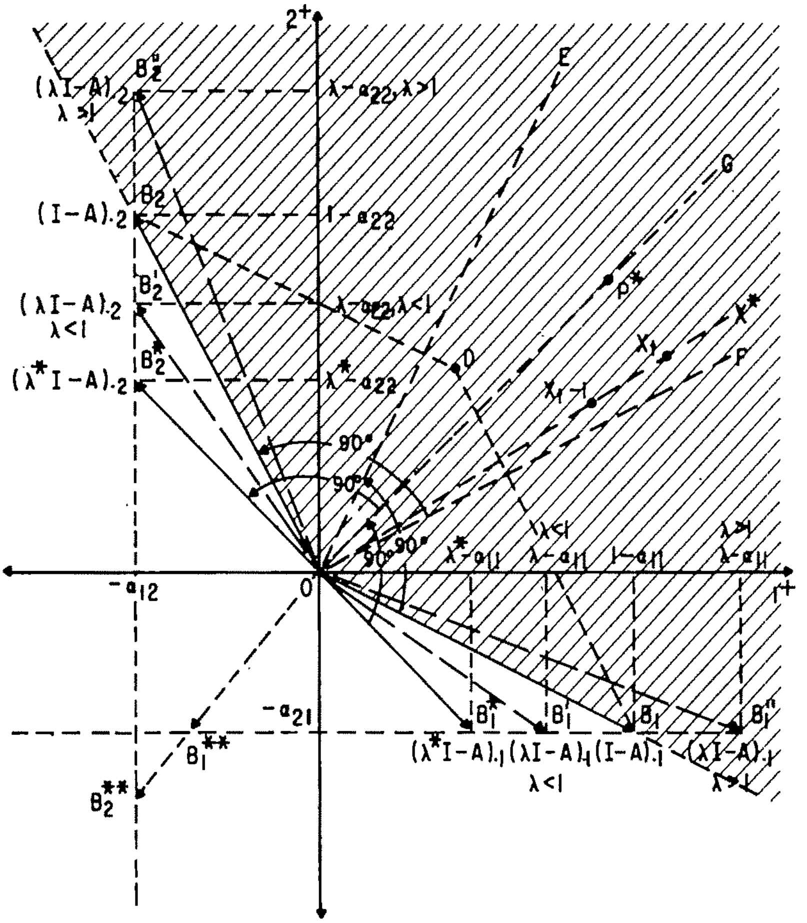


FIGURE 1. THE CASE OF VIABLE SOLUTIONS WITH TWO GOODS

$$\{y \in R^n \mid y = (I - A)x,\ x \geqq 0\} \supset \{y \in R^n \mid y > 0\}$$

Condition 1) states that the cone generated by $[I-A]$ contains at least one element from the positive orthant. Since this cone either contains all of the positive orthant or none of it, condition 1) is equivalent to condition 8). Condition 2) states that the cone generated by $(I-A)$ contains all of the nonnegative orthant (the positive orthant together with the positive axes) and is thus obviously equivalent to condition 8).

Condition 3) states that there exists a positive vector p that forms a strictly positive scalar product with every column of $[I-A]$. This simply means that p forms an acute angle with every column of $[I-A]$. To say that every column of $[I-A]$ forms an acute angle with a strictly positive vector is clearly equivalent, moreover, to the statement that the columns of $[I-A]$ bend toward the positive orthant or the statement that the cone generated by the columns of $[I-A]$ contains the positive orthant. Thus condition 3) is equivalent to condition 8). Notice in this regard that the interior of the cone *EOF* in Figure 1 represents the set of price

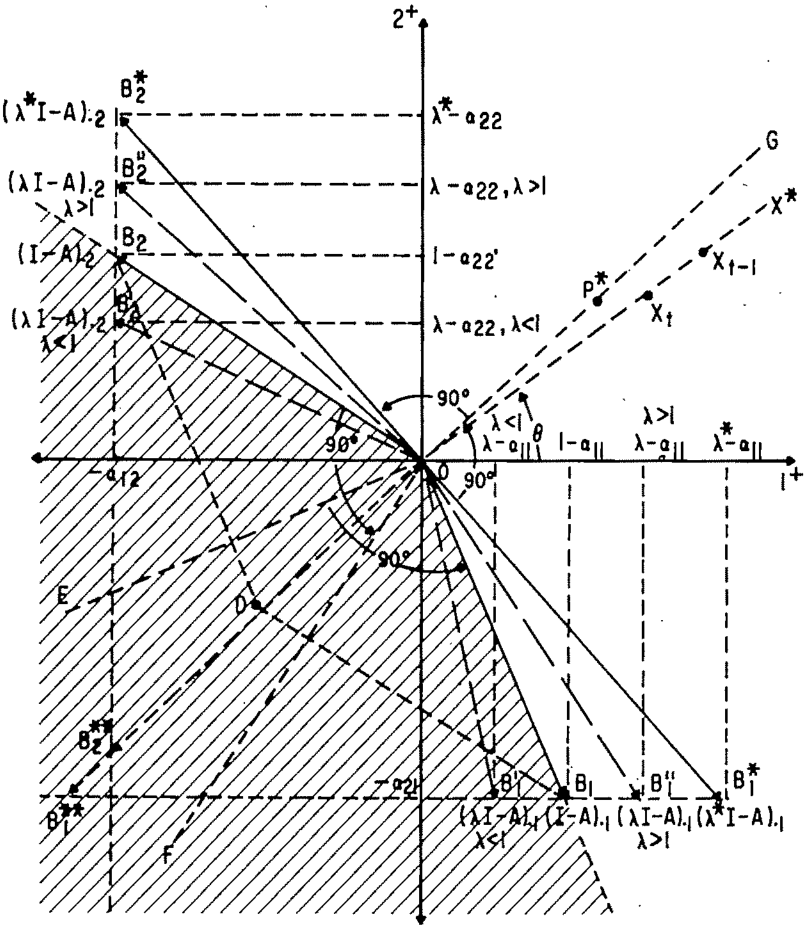


FIGURE 2. THE CASE OF NO VIABLE SOLUTIONS WITH TWO GOODS

vectors at which both industries could operate profitably. (The analogous cone *EOF* in Figure 2, on the other hand, contains only negative prices.)

Condition 4) states that there exist x_t, x_{t-1} such that $x_t > x_{t-1}$ and $Ax_t \leqq x_{t-1}$. If we define

$$c = x_t - x_{t-1} > 0$$

we may simplify this condition as follows:

$$x_{t-1} - Ax_t \geqq 0$$

$$x_{t-1} - A(x_{t-1} + c) \geqq 0$$

$$x_{t-1} - Ax_{t-1} \geqq Ac > 0$$

$$(I - A)x_{t-1} > 0$$

Condition 4) states, therefore, that the cone generated by the columns of $[I-A]$ contains an element from the positive orthant. Thus condition 4) is equivalent to condition 1) which has already been shown to be equivalent to condition 8).

Condition 5) is the most difficult to interpret. For the case of two goods ($n=2$) it requires that

$$\det (1 - a_{11}) > 0,$$

$$\det \begin{bmatrix} 1 - a_{11} & -a_{12} \\ -a_{21} & 1 - a_{22} \end{bmatrix}$$

$$= (1 - a_{11})(1 - a_{22}) - a_{12}a_{21} > 0$$

Notice that it is necessary for the second-order determinant to be positive that $(1-a_{22})$ be positive also. A moment's consideration reveals that if either $(1-a_{11})$ or $(1-a_{22})$ were negative the corresponding column of $(I-A)$ would lie in the negative (third) quadrant and it would be impossible for the cone generated by the columns of $[I-A]$ to contain the positive quadrant irrespective of the position of the other column. This reasoning is also correct for the case of three or more goods.

Sufficient conditions for higher order determinants to be positive are more troublesome. Routine calculation reveals, however, that the determinant of $[I-A]$ for the case of two goods is (except for sign) the area of the parallelogram determined by the two columns of $[I-A]$.[12] That is, in Figure 1 the determinant of $[I-A]$ is equal to the area of parallelogram B_1OB_2D and in Figure 2 the determinant of $[I-A]$ is equal to ***minus*** the area B_1OB_2D. The sign in each case depends upon whether the area is obtained by "sweeping" counterclockwise from B_1 to B_2 (positive sign) or by sweeping clockwise from B_1 to B_2 (negative sign).[13] Thus requiring that both determinants be positive is equivalent to requiring that the cone generated by the (two) columns of $[I-A]$ contain the positive quadrant.

The case of three goods is only slightly more difficult. Here the determinant of $[I-A]$ can be interpreted (except for sign) as the volume of the parallelepiped formed by the (three) columns of $[I-A]$.[14] (This parallelepiped is illustrated in Figure 3.) The second-order leading principal minor is moreover (except for sign) area of the parallelogram obtained by p jecting $(I-A)_{.1}0(I-A)_{.2}D$ onto the pla formed by the first and second coordin axes.

This second-order leading principal nor is positive if the direction of movem from $(I-A)_{.1}$ to $(I-A)_{.2}$ is countercloc wise in the plane formed by the first ar second coordinate axes.[15] If the first- a second-order leading principal minors a positive the sign of the determinant $[I-A]$ depends upon whether the dire tion of movement from $(I-A)_{.2}$ $(I-A)_{.3}$ is counterclockwise (positi sign) or clockwise (negative sign) in th plane formed by the second and thir coordinate axes.

The three goods case with all leading principal minors positive is illustrated in Figure 3. That the first, second, and third-order leading principal minors are positive is illustrated, respectively, by the following observations: $(1-a_{11})$ is positive, the direction of sweep is counterclockwise from $(I-A)_{.1}$ to $(I-A)_{.2}$, and the directions of sweep from $(I-A)_{.1}$ to $(I-A)_{.2}$ and from $(I-A)_{.2}$ to $(I-A)_{.3}$ are both counterclockwise.[16] These conditions are clearly equivalent to the statement that the cone generated by $[I-A]$ contains the positive orthant.[17]

---

[12] If $[I-A]$ were singular, the parallelogram would degenerate (collapse) into a line segment which has zero area.

[13] The order "B_1 to B_2" is crucial. Recall that switching columns changes the sign of the determinant.

[14] If $[I-A]$ were singular the parallelepiped would degenerate (collapse) into a parallelogram which has zero volume. See Michael Spivak, p. 83, for a general discussion of the determinant as an "oriented" volume.

[15] There is a simple "rule of thumb" for determining the direction of counterclockwise sweep. Point the thumb of your *right* hand along the positive direction of one axis and the fingers curl in the direction of counterclockwise sweep (positive orientation) for movements in the perpendicular plane.

[16] The sign of the determinant of $[I-A]$ is determined in general by interchanging columns of $[I-A]$ until the resulting matrix has a ***normal orientation*** (until every column lies in the direction of counterclockwise sweep from the preceding column). If the number of required "switches" is odd, then the sign of the original determinant is negative; otherwise, the sign is positive. Note that in Figure 2 $[I-A]$ has a normal orientation to begin with and thus requires zero switches.

[17] It should be noted that if the *leading* principal minors are all positive, then *all* principal minors must be positive. In Figure 2, for example, the remaining

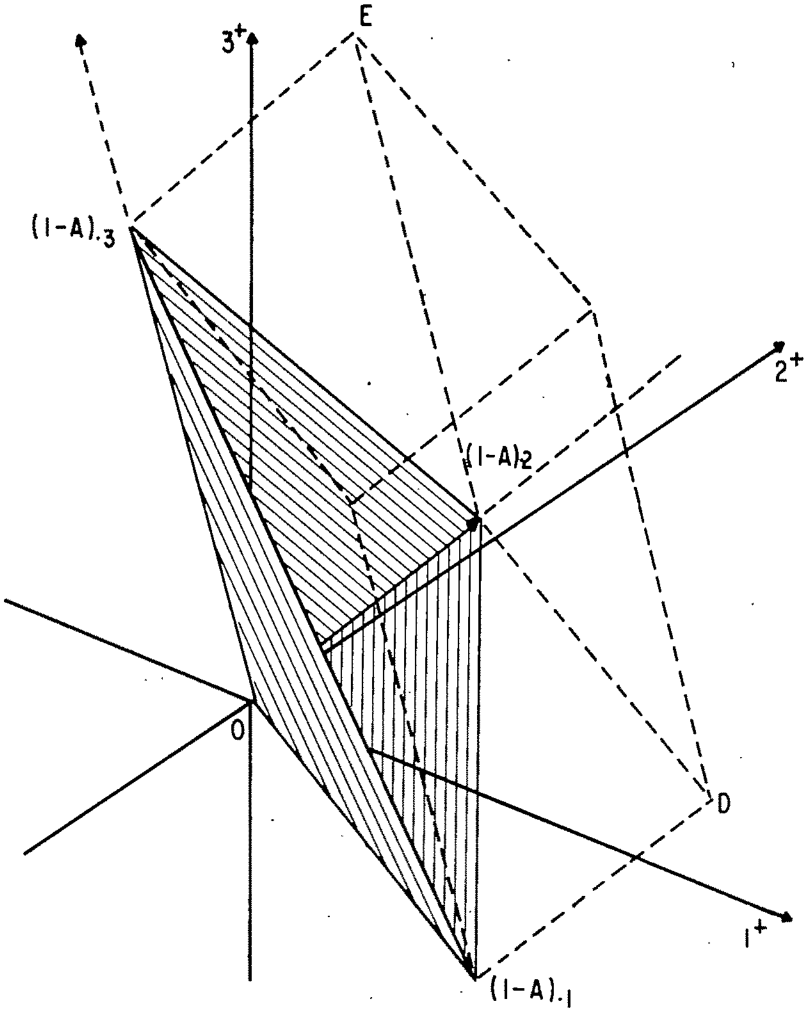


FIGURE 3. THE VIABLE SOLUTIONS CASE WITH THREE GOODS

Similar reasoning (with hyper-parallelepipeds) would establish the same result for an arbitrary number of goods and we may conclude that condition 5) is equivalent to condition 8).

Condition 6) states that there exists a vector

$$d = (d_1, \ldots, d_n) > 0$$

such that $d(I-A)>0$.

This condition is obviously equivalent to condition 3) which has already been discussed.

Condition 7) requires that the largest real root of A be less than one. To see this, first consider the case of two goods illustrated in Figure 1. The dominant root is

---

irst-order principal minors $(1-a_{22})$ and $(1-a_{33})$ must be positive as well as the second-order principal minors ormed by the second and third rows and columns and by the first and third rows and columns of $[I-A]$. The only trick in seeing this is to realize that in the last case he relevant sweep is from the third column to the first ather than from the first to the third. (The column that "preceeds" the first column in the cyclic order 1 2 3 1 . . . is the third).

the largest real number λ such that[18]

$$\det\begin{bmatrix}\lambda - a_{11} & -a_{12}\\ -a_{21} & \lambda - a_{22}\end{bmatrix} = 0$$

Notice that for $\lambda=1$ the determinant is equal to the area B_1OB_2D (in Figure 1) which is certainly not zero. It cannot be the case, therefore, that $\lambda=1$ is the dominant root.

Suppose now that we increase the "trial" value of λ to some value greater than one. This has the effect of bending the columns of $[\lambda I-A]$ further toward the positive quadrant and is illustrated in Figure 1 by the shifts from OB_1 to OB_1'' and from OB_2 to OB_2''. It is apparent that further increasing the trial value of λ will never make the determinant of $[\lambda I-A]$ equal to zero. The only case in which the determinant is zero occurs when both columns of $[\lambda I-A]$ lie in the same line (when the parallelogram collapses into a line segment). Increasing λ above 1 in Figure 1 simply bends these columns the wrong way.

Since any value of λ above one cannot be the dominant root, consider the effect of reducing λ below one. This has the effect of bending the columns of $[\lambda I-A]$ away from the positive quadrant and toward the negative quadrant and is illustrated by the shifts from OB_1 to OB_1' and from OB_2 to OB_2'. In fact, when λ is zero both columns of $[\lambda I-A]$ belong to the negative quadrant. It must be the case, therefore, that at some value of λ between one and zero the columns of $[\lambda I-A]$ cease to bend toward the positive quadrant and begin to bend toward the negative quadrant. At this point the columns of $[\lambda I-A]$ lie in the same line—$B_1^*OB_2^*$ in Figure 1. This value of λ is, of course, the dominant root and is denoted by λ^*. That is, λ^* appears at the point where the dominant diagonal vanishes.

Similar reasoning applied to Figure 2 leads to the conclusion that the dominant root must be greater than one when the cone contains the negative quadrant. The only other possibility, that $[I-A]$ is singular, corresponds to a dominant root equal to one. It is clear, therefore, that in the case of two goods the dominant root of A being less than one is equivalent to the cone generated by the columns of $[I-A]$ containing the positive quadrant.

For the case of three goods the reasoning is quite analogous. Here the problem is to choose λ so that the columns of $[\lambda I-A]$ lie in a plane. If the cone generated by the columns of $(I-A)$ contains the positive orthant, we must choose λ less than one in order to bend these columns away from the positive orthant and toward the negative orthant (see Figure 3). If the columns of $[I-A]$ generate a plane, then $[I-A]$ is singular and the dominant root is one. Lastly, if the columns of $[I-A]$ generate a cone which contains the negative orthant, then we must choose λ greater than one in order to bend these columns away from the negative orthant and toward the positive orthant. It is clear that this reasoning holds for an arbitrary number of goods and we may conclude that condition 7) is equivalent to condition 8).

An indirect implication of the above analysis is that the dominant root is nonnegative in *all* cases—Theorem (a). This can be seen as follows. For *any* negative λ the columns of $[\lambda I-A]$ belong to the nonpositive orthant.[19] On the other hand, by making λ sufficiently positive $[\lambda I-A]$ can be made to generate a cone which contains the positive orthant (each column of $\lambda I-A]$ approaches the respective positive coordinate axis as λ becomes larger and larger). There must exist a real characteristic root, therefore, between the negative and positive values of λ (which corresponds to the point at which the columns of $[\lambda I-A]$ bend neither toward the negative nor the positive orthants). Since this is true for any negative λ this root must be nonnegative. That is, there must exist λ^* greater than or equal to zero which satisfies these conditions.[20]

---

[18] λ is a solution of $Ax=\lambda x$ iff det $[\lambda I-A]=0$. See Hadley.

[19] The nonpositive orthant is the set $\{y\in R^n \mid y\leq 0\}$.

Notice that the columns of $[\lambda^* I-A]$ generate a degenerate or flat cone (a linear subspace). (This cone is illustrated by the infinite line through $B_2^*OB_1^*$ in Figure 1 or Figure 2.) The subset of this cone that would be obtained by considering only those nonnegative weights for the columns of $[\lambda^* I-A]$ that sum to one, i.e.,

$$\left\{y \mid y = \sum_{i=1}^{n} \alpha_i(\lambda^* I - A)_{\cdot i}, \quad \alpha_i \geqq 0, \ \sum_{i=1}^{n} \alpha_i = 1\right\}$$

is a convex polyhedral set whose vertices are formed by the columns of $[\lambda^* I-A]$ and which contains the origin. (This subset is illustrated by the line segment $B_1^*OB_2^*$ in Figure 1 or Figure 2 and would be a triangle passing through the origin in Figure 3.) That set of weights x_i^* which generates the origin (causes the sum of the weighted columns to be the origin) is a semipositive vector which satisfies

$$\sum_{i=1}^{n} x_i^*(\lambda^* I - A)_{\cdot i} = 0$$

$$(\lambda^* I - A)x^* = 0$$

$$Ax^* = \lambda^* x^*$$

and is thus the semipositive characteristic vector associated with the dominant root —Theorem (b).[21]

Theorem (c) follows from the fact that increasing *any* element of A either has no effect upon the (degenerate) cone generated by the columns of $[\lambda^* I-A]$ (possible only if A is decomposable) or it causes these columns to bend toward the negative orthant. For example, increasing either a_{11} or a_{21} in Figure 1 or Figure 2 rotates the first column of $(\lambda^* I-A)$ toward the negative quadrant. To bend these columns back toward the positive orthant (to obtain the dominant root of the new matrix) requires, therefore, increasing λ above the value of the original dominant root.

Theorems (d) and (f) both involve the fact that increasing λ above λ^* causes the columns of $[\lambda I-A]$ to bend toward the positive orthant. This means that λ being greater than λ^* is equivalent to the cone generated by the columns of $[\lambda I-A]$ containing the positive orthant. The discussion of condition 5) revealed, however, that the latter condition is necessary and sufficient for all leading principal minors of $[\lambda I-A]$ to be positive. Thus λ being greater than λ^* is equivalent to all leading principal minors of $[\lambda I-A]$ being positive —Theorem (f).

Theorem (d) says, in essence, that the inverse of $[\lambda I-A]$ is semipositive if and only if the cone generated by the columns of $[\lambda I-A]$ contains the positive orthant. To see this, let us construct the inverse of such a matrix. Notice that in Figure 1 the columns of $[I-A]$ generate a cone which contains the positive orthant. We may determine the rows of $[I-A]^{-1}$ by the following method. The fact that

$$[I - A]^{-1}[I - A] = I$$

Means that the first *row* of $[I-A]^{-1}$ must form an acute angle with the first ***column*** of $[I-A]$ (the product of these vectors must be one) and a right angle with the second *column* of $[I-A]$ (the product of these vectors must be zero). The first *row* of $[I-A]^{-1}$ must lie, therefore, somewhere along the extended line OF in Figure 1.

---

[20] The fact that λ^* is strictly positive when A is indecomposable is suggested by Figures 1, 2, and 3. For λ^* actually to be zero it is necessary that A be decomposable and singular.

[21] In Figure 1, notice that B_1^* and $(|B_1^*|/|B_2^*|)B_2^*$ have the same length. Thus $B_1^*+(|B_1^*|/|B_2^*|)B_2^*=0$ and $x^*=(1, |B_1^*|/|B_2^*|)>0$.

Similar reasoning reveals that the second row of $[I-A]^{-1}$ must lie along OE. Notice that this inverse matrix must be positive (or, perhaps, semipositive if A were decomposable).[22]

For an arbitrary number of goods one obtains a *row* of $[\lambda I-A]^{-1}$ by finding a vector that forms a zero scalar product (a right angle) with all *columns* of $[\lambda I-A]$ save one and a positive scalar product (an acute angle) with that one. Thus every row of the inverse must form a nonnegative scalar product (a nonobtuse angle) with every vector in the cone spanned by the columns of $[\lambda I-A]$. Since the cone spanned by the columns of $[\lambda I-A]$ contains the semipositive orthant if and only if λ is greater than λ^*, it follows that *every* row of the inverse must form a nonnegative scalar product (a nonobtuse angle) with every semipositive vector—that every row of the inverse is semipositive—if and only if λ is greater than λ^*. Thus for any number of goods, $[\lambda I-A]^{-1}$ being semipositive is equivalent to λ being greater than λ^*—Theorem (d).

To gain insight into the issue of decomposability consider the case in which the columns of $[I-A]$ span the positive orthant. The fact that

$$x = [I - A]^{-1}c$$

where x is the vector of gross output levels required to satisfy the vector of final demands c, means that

$$\frac{\partial x_i}{\partial c_j} = [I - A]_{ij}^{-1} \qquad i, j = 1, \ldots, n$$

may be interpreted as the increase in the gross output of the ith industry required to satisfy an increment in the final demand for the jth good. Now, if A is decomposable then there exists a partition of $\{1, 2, 3, \ldots, n\}$ into nonempty sets I and J such that[23] $a_{ij}=0$, $i\in I, j\in J$. This means that the columns of $[I-A]$ corresponding to the indices in J have zeros in the rows corresponding to the indices in I or simply that the J columns of $[I-A]$ lie in a linear subspace formed by the J positive coordinate axes. The I rows of $[I-A]^{-1}$ must be perpendicular (orthogonal) to each of the J columns of $[I-A]$, moreover. This means that the I rows of $[I-A]^{-1}$ must belong to the linear subspace formed by the I coordinate axes or that the I rows of $[I-A]^{-1}$ must have zeros in the columns corresponding to the indices in J. Thus

$$[I - A]_{ij}^{-1} = 0 \qquad i \in I, j \in J$$

and no change is required in the gross outputs of any of the I industries to increase final output in any of the J industries.

An example should clarify this. Consider the case of three goods where $[I-A]$ has the sign pattern:

$$\begin{bmatrix} + & - & - \\ - & + & - \\ 0 & 0 & + \end{bmatrix}$$

Clearly A is decomposable and $I=\{3\}$, $J=\{1, 2\}$ form the required partition. Notice that the first and second columns (the J columns) of $[I-A]$ lie in the plane (the linear subspace) formed by the first and second coordinate axes. For the third row of $[I-A]^{-1}$ to be perpendicular to both the first and second columns of $[I-A]$ it must be perpendicular to the plane determined by the first and second coordinate axes. The only possibility is that the third row (the I rows) of $[I-A]^{-1}$ must lie along the third coordinate axis

---

[22] If a_{21} were zero, for example, the first column of $[I-A]$ would lie along the positive first axis and the second row of $[I-A]^{-1}$ would lie along the positive second axis (and the lower left-hand corner of the inverse is zero). The partition of $\{1, 2\}$ is given in this case by $I=\{2\}$, $J=\{1\}$.

[23] Recall that I and J partition $\{1, 2, \ldots n\}$ iff $I\cap J=\phi$ and $I\cup J=\{1, 2, \ldots n\}$.

(the linear subspace formed by the I coordinate axes). This third row must, therefore, have zeros in the first and second places, i.e.,

$$[I - A]^{-1}_{31} = [I - A]^{-1}_{32} = 0$$

Theorem (e) requires some consideration of the other characteristic roots and vectors of A.[24] For the case of two goods illustrated in Figure 1 or Figure 2, the other root corresponds to the point at which $(\lambda I-A)_{\cdot 1}$ and $(\lambda I-A)_{\cdot 2}$ "cross" in the negative quadrant. This crossing must occur since $(\lambda I-A)_{\cdot 1}$ rotates clockwise through the negative quadrant toward the negative first axis as λ becomes smaller and smaller (more and more negative) and $(\lambda I-A)_{\cdot 2}$ rotates counterclockwise through the negative quadrant toward the negative second axis as λ becomes smaller and smaller. This crossing is illustrated by the vectors B_1^{**} and B_2^{**} in Figures 1 and 2. Notice that the characteristic vector corresponding to this root (the weights which make the columns B_1^{**} and B_2^{**} sum to the origin) cannot be semipositive since the origin does not belong to the line segment $B_1^{**}B_2^{**}$.

The case of three goods is quite similar. As λ becomes smaller and smaller, $(\lambda I-A)_{\cdot 1}$ (in Figure 3) rotates through the negative orthant toward the negative first axis, $(\lambda I-A)_{\cdot 2}$ rotates through the negative orthant toward the negative second axis, and $(\lambda I-A)_{\cdot 3}$ rotates through the negative orthant toward the negative third axis. Either there are two crossings of the type that one vector crosses the plane determined by the other two (the other two roots are real and distinct), or one vector coincides with the plane determined by the other two but does not actually cross it (two repeated real roots), or there are no crossings at all (two conjugate complex roots).[25] Any other real characteristic roots and vectors must correspond, therefore, to real crossings in the negative orthant. This means that the convex polyhedral set whose vertices are determined by the columns of $[\lambda I-A]$ (when evaluated at any other real root) cannot contain the origin. Any other real characteristic vector (the weights which make these columns sum to the origin) cannot, therefore, be semipositive. This reasoning holds for an arbitrary number of goods and we may conclude that x^* is the only semipositive characteristic vector—Theorem (e).

It should be noted in passing that λ^* and x^* have an economic interpretation. The search for balanced growth

$$x_t = \sigma x_{t-1}$$

with full employment

$$Ax_t = x_{t-1}$$

leads to the realization that $\sigma = 1/\lambda^*$ is the maximum balanced growth factor and x^* represents the balanced growth proportions. This means that $1/\lambda^* - 1$ can be regarded as the maximum (percentage) growth rate.

Consider, alternatively, the problem of determining the profit rate of the jth industry when prices are given by the (row) vector p. This profit rate is obtained by finding the discount rate, r, which makes the present value of receipts equal to the present value of production costs, i.e., the solution r to

$$p_j/(1 + r) - pA_{\cdot j} = 0$$

One naturally wonders if there exists an equilibrium vector of prices such that the profit rate is the same in all industries.

---

[24] There are a total of n (not necessarily distinct) characteristic roots and vectors of A. See Hadley.

[25] The latter possibility is due to the fact that the plane formed by any two vectors rotates as it sweeps through the negative orthant. It is therefore possible for this plane to "roll over" the other vector without ever actually crossing it.

That is, is there a set of prices p and a common profit rate r that satisfies

$$p/(1+r) - pA = 0$$

or, more simply,

$$p[1/(1+r)I - A] = 0$$

Clearly, $1/(1+r)=\lambda^*$ or $r=1/\lambda^*-1$ is the required common profit rate and $p=p^*$ (p^* is the characteristic *row* vector associated with the dominant root λ^*) is the required equilibrium price vector.

The vector p^* can be obtained geometrically by noticing that p^* must form a right angle with every column of $[\lambda^* I - A]$. For the case of two goods illustrated in Figure 1 or Figure 2, p^* must lie along line OG. Notice that p^* is positive in this case and that in general the same propositions can be established for this dominant row vector as were established for the dominant column vector x^*.

Finally notice that if λ^* is less than one, the maximum growth and profit rate, $1/\lambda^*-1$, is positive while if λ^* is greater than one, the maximum growth and profit rate is negative (see Figures 1 and 2).

REFERENCES

E. Burmeister and A. R. Dobell, *Mathematical Theories of Economic Growth,* New York 1970.

G. Hadley, *Linear Algebra,* Reading, Mass. 1961.

H. Nikaido, *Introduction to Sets and Mappings in Modern Economics,* Amsterdam 1970.

M. Spivak, *Calculus on Manifolds,* New York 1965.

Toward an Analysis of Postdisaster Cooperation

By Louis De Alessi*

There is evidence that, in the period immediately following a disaster, cooperative behavior increases.[1] Individuals apparently become more generous toward each other, offering goods and services at prices below those they could have obtained under the existing time-dated pattern of current and expected demand, and supply conditions. Thus, for example, the prices of such things as shelter and "emergency" foods often do not increase enough to clear the market, and the allocation of these and other goods is marked by a variety of formal and informal nonprice rationing schemes.[2]

A number of apparently conflicting hypotheses have been offered to explain this behavior. Accordingly, it seems useful to reexamine the problem and to place current contributions in their proper perspective.

Section I contains a summary of the arguments regarding whether disasters occasion a shift in individual utility functions toward more charity, as suggested by Douglas Dacy and Howard Kunreuther, or simply a decrease in the price of charity, as suggested by the author. Section II contains a review of Jack Hirshleifer's addendum that some postdisaster cooperation is due to alliance activities.[3] Section III contains a critique of Christopher Douty's attempt, based upon the alliance hypothesis, to rule out utility interdependence entirely and to explain postdisaster altruism essentially in terms of informal insurance arrangements. The range of application of the various hypotheses is examined further in Section IV, and Section V contains a few concluding remarks.

I. Utility Interdependence

Dacy and Kunreuther, in studying the aftermath of the Alaskan earthquake of 1964, were struck by the apparent increase in cooperative behavior. Having noted that other disasters had given rise to similar results, they conjectured that disasters yield short-run structural changes in individual utility functions toward more "community feeling." Individuals presumably derive utility from increases in the welfare of others,[4] and respond to the suffering and distress occasioned by a disaster by attaching greater weights to the welfare of the victims. Utility functions thus shift, and individuals behave with greater charity toward the victims. In due course, as the emergency wanes, utility functions recede toward their original positions.

---

* Professor of economics, George Washington University. I am indebted to the National Science Foundation for financial assistance and to Jack Hirshleifer, Douglas W. Webbink, and an anonymous referee for helpful comments.

[1] A disaster is defined as the sudden destruction of physical assets. Charity is defined as the intentional transfer of resources below their market prices for the purpose of making the recipients better off, as the donors see it, and is associated with utility interdependence. Cooperation is defined as charitable and seemingly charitable behavior (e.g., it includes wealth-maximizing business "gifts") and is used synonymously with altruism; thus, it excludes normal market interactions which in a broader context are a prime example of cooperative activity.

[2] G. Sjoberg, Douglas Dacy and Howard Kunreuther, Jack Hirshleifer, Christopher Douty, and the author, among others, discuss some of the pertinent evidence and note additional sources of information.

[3] A few words regarding the actual chronology of these contributions may be useful to readers puzzled by the apparent inconsistency of the publication dates shown. The study by Dacy and Kunreuther (1969) first appeared during 1966–67 in a series of Internal Notes, approved for external circulation, published by the Institute for Defense Analyses, Arlington, Virginia. The two papers by the author (1967, 1968) were drafted during the second half of 1966 as a result of several enjoyable conversations with Dacy and Kunreuther. One of the papers (1968) initially appeared in 1966 as an Internal Note, approved for external circulation, published by the Institute for Defense Analyses. Hirshleifer footnoted most of this material in his paper (1967), which he circulated privately.

I have argued that a shift in utility functions is not necessary to explain the behavior observed. If individual utility functions are initially interdependent, then a disaster lowers the welfare of the victims and enables donors to obtain more utility from any given gift.[5] The search and information costs of finding worthy recipients (as the donors see it) also decrease, and the opportunity to make others better off expands dramatically. In effect, a disaster lowers the price of charity. Individuals predictably engage in more charitable behavior,[6] reallocating some of their resources from other activities (including other charities) to the relief of disaster victims until each donor's utility-maximizing equilibrium is restored.[7] Since a shift in individual utility functions is not observable and yields the same implications, it is not necessary to explain the phenomena observed and thus may be rejected by the rule of Occam's razor.

The observation that, following a disaster, some business firms sell some goods and services at prices below the market-clearing level similarly is not inconsistent with an unchanged pattern of tastes. The owners' wealth-maximization hypothesis already implies that the managers of certain firms have the incentive to allocate business resources to altruism in order to acquire goodwill or to avoid badwill (see Orace Johnson). A disaster clearly provides additional opportunity for such behavior. Moreover, if the structure of property rights provides business decision makers with the opportunity to increase their own welfare at the expense of the fisc or of owners' wealth, then a decrease in the price of charity induced by a disaster implies that more business resources will be allocated to charity than otherwise would have been the case (see the author, 1968).

II. Alliance Hypothesis

Hirshleifer concurred with much of my analysis,[8] but suggested that it failed to account for some of the behavior observed. To remedy the situation, he proposed to extend the alliance hypothesis of Mancur Olson and Richard Zeckhauser to post-disaster situations.

---

[4] More precisely, the inputs used by individuals in satisfying the goals which enter their utility functions include some goods and services consumed by others. "Others" may consist of only one or two relatives and friends, or may include larger groups distinguished by some characteristic (e.g., religion, race) relevant to the chooser. Benevolence implies that individuals make gifts, including tied gifts, and there is much evidence of this behavior. For some individuals, however, an increase in the consumption of certain commodities (e.g., alcoholic beverages, drugs) by others may be a source of disutility; some of the consequences of malevolence will be noted later on.

[5] See the author (1967) for a diagrammatic exposition of the two contrasting viewpoints. As usual, individual utility functions are hypothesized to be single-valued, convex, and twice differentiable.

[6] The relative price of charity falls for all individuals, including victims. In the case of some victims, the movement down along the demand curve (the price effect) may be more than offset by the decrease in the demand curve occasioned by the disaster-caused loss of wealth, yielding a reduction in the amount of charity undertaken. Charity is taken throughout to be a normal or superior good.

[7] The utility interdependence hypothesis implies that aid will decrease and eventually cease as victims recover. Following a disaster, the victims—aided by gifts and by the resumption of community activities—begin to recover and to revise the profile of their assets and liabilities to cope with the new situation. As the welfare of the victims increases, donors receive less utility from additional gifts, and aid to victims decreases and eventually stops.

[8] Thus, Hirshleifer noted that "If [community feeling] means that disaster brings about a shift in the utility function, the phenomenon lies in the realm of psychology—so there is little for the economist to discuss" (p. 4).

Briefly, the alliance hypothesis asserts that the continuation of organized society is a collective good which benefits most of society's members (see Hirshleifer, p. 6). This provides even narrowly selfish individuals with the incentive to engage in alliance-supportive activities. The duties assumed under the social contract are enforced by custom and policemen (pp. 5–7). Disasters enhance the marginal effectiveness of individual actions for survival of the alliance, and thereby give rise to good behavior. The greater the likelihood that the society will be overpowered by the disaster, however, the greater the probability that bad behavior will occur, as each individual increasingly looks out for his own narrow interests. If stress exceeds whatever point may be optimal for a society in view of its customs, culture, and size, then the alliance will collapse (pp. 8–9).

In preparing his case for the alliance hypothesis, Hirshleifer agrees that the disaster-induced increase in the relative marginal utility to an individual of victims' consumption explains " . . . why relatively better-off individuals in disasters tend to help their less fortunate neighbors" (p. 4) and " . . . the universally observed fact that relief . . . tends to flow into a damaged area from the untouched outside" (pp. 4–5). On the other hand, he finds that:

> The opposite side of the coin, the implication that those relatively worse-off tend to be less altruistic than before, is also plausible, but there seems to be no evidence relating degree of altruistic activities inversely to the degree of relative deprivation. Also, this view provides no explanation as to why *general* impoverishment is met with general "good behavior," nor as to why exceptional cases occur in which "bad behavior" is observed. [p. 5]

The behavior at issue, however, does not appear to be inconsistent with the utility interdependence hypothesis. First, lack of evidence about an implication does not provide sufficient ground for doubting its validity. If the hypothesis suggested by Hirshleifer has been deduced correctly from the axioms of consumer theory under the appropriate antecedent conditions, then it receives indirect support from the evidence pertaining to other aspects of consumer choice and it may be taken to hold until more direct evidence becomes available.[9]

More importantly, Hirshleifer is quite right in calling attention to the possibility of bad behavior following a disaster. Such behavior, however, is explicable within the existing analytical framework. *Ceteris paribus*, the decrease in the price of charity faced by an individual within the disaster area is offset—to the extent that he is also a victim—by his loss of wealth. As a result, the degree of altruism observed within a community will depend upon the extent and dispersion of the damage among its members.[10] If the disaster is localized and the individual's survival is not at stake (conditions which were implicit in previous discussions), then cooperative behavior may be expected to dominate. More tragic conditions, however, generate both lower prices of charity and greater losses of wealth. Beyond some point, the wealth effect will begin to dominate, yielding a lower level of cooperative behavior. As the seriousness of the disaster increases, therefore, not only are more individuals harmed more seriously, but the decrease in altruism implies that a victim's expectation of *receiving* aid is also reduced. *Ceteris paribus*, this will increase the relative benefits

---

[9] A rigorous test of the hypothesis in question would be helpful. In doing so, however, it would be necessary to take into account such things as changes in the enforcement of laws and differences in the relative price of charity confronting different individuals.

[10] There is also the problem of weighing acts of altruism under alternative conditions, such as a helping hand when survival is at stake relative to a gift of shelter under less trying circumstances.

(for example, survival) to be obtained from bad behavior, and more of the latter will occur. A weakening of the enforcement mechanism similarly will encourage bad behavior.[11] Altogether, it is certainly plausible that the magnitude and form of some disasters will be such that bad behavior will dominate. In general, the mix of good and bad behavior and which one will dominate will depend upon such things as the relative taste for these activities (as influenced by custom and culture), the change in their relative prices, the extent of the disaster, and the size and kinship of the group.

In justifying the case for alliance, Hirshleifer denies " . . . the premise that substantially everyone's utility map has a positive partial derivative with respect to 'other fellows' income" (p. 5). If the statement is intended to deny that substantially everyone's income enters substantially everyone else's utility function, then Hirshleifer is quite right. Such a premise, however, is not necessary to the analysis. A sufficient condition is that substantially everyone's utility function has a positive partial derivative with respect to at least *some* other individuals' income, say just a few friends and relatives. This premise seems plausible, but even weaker ones would work. The validity of economic propositions does not require that all or even substantially all actors respond to a change in circumstances. In the case of disasters, where friends and relatives living in the same community are affected, even a fairly selective response at the margin would be enough to yield the consequences predicted. And some individuals undoubtedly recognize a larger kinship group.

Having explicitly adopted a weaker premise of utility interdependence, Hirshleifer suggests that many cases of negative taste for altruism—at least at the margin—may be expected to give rise to conflict and to cloud the effect of general impoverishment upon behavior. The importance of this point, however, could easily be overstated since the disaster-induced consequences of positive and negative altruism are not wholly symmetric. Using Hirshleifer's example (p. 5), let A view keeping up with the Joneses as a problem and their income as a bad. If the Joneses become victims of a disaster and A is untouched, A will gloat and move to a higher level of utility. As long as laws are enforced, however, A will find it costly, just as he did before the disaster, to visit additional misfortunes upon the Joneses. In general, A will simply abstain from aiding the Joneses while assisting the Smiths, who may be poorer, friendlier, or whatever. Indeed, if the Joneses are hurt enough, A may even help them. None of these statements, of course, is intended to deny the possibility of bad behavior (including the settling of old scores) noted earlier.

Although the alliance hypothesis does not appear to be necessary to explain the behavior observed, its case is appealing. Individuals do form organized societies in order to reap the benefits of increased social interaction (including greater specialization), and this implies that they have the incentive to preserve the society. Nevertheless, even if disasters and other threats multiply manifold the marginal effectiveness of individual actions in preserving society (see Hirshleifer, p. 7), an individual would undertake such actions voluntarily only to the extent that *his* benefits exceeded *his* costs. Otherwise, he would take a free ride. And since the social contract at best is vague, the enforcement of alleged duties is difficult and thus presents individuals with considerable scope for free riding.[12] Custom may help, and

---

[11] Major disasters often are accompanied by the introduction of martial law, a step designed to discourage bad behavior by imposing higher penalties earlier in time.

[12] Recall that the problem is to explain *voluntary* cooperation. Although policemen seek to enforce society's

Hirshleifer notes that the social contract is partially self-enforcing as education is used to convert duties into customary behavior (p. 6). However, on this tack, alliance seems to explain why the welfare of others is a good. The view that some utility interdependence is a desirable survival characteristic is encouraged by the observation of altruistic behavior among primates and other animals (see Dennis Krebs, pp. 261–62 and Konrad Lorenz), and would help to explain why, on the postdisaster evidence available, alliance behavior seems to be indistinguishable from utility interdependence.[13] Hirshleifer does not suggest tests which could be used to discriminate rigorously between the two alternatives, and such tests—for the reasons noted earlier—seem difficult to formulate.[14]

III. Informal Insurance

Douty sought to establish that postdisaster cooperative behavior may be wholly explained by a narrowly conceived view of self-interest, one which excludes utility interdependence entirely (p. 583).[15] Basing his case upon an expanded version of Hirshleifer's alliance hypothesis, he proposed informal insurance as the key explanatory hypothesis.

According to Douty, consensual agreement[16] provides the institutional environment which specifies the range of actions an individual can undertake in his own behalf and permits market interdependence and specialization (p. 584). A major task of government is to make sure that the benefits of stable institutions are widely diffused within a society (p. 584).

A disaster initially fragments a community into small kinship groups of family and close associates. As soon as the individual has been assured of the survival of members of his kinship group, however, he becomes concerned with the survival of the larger community and with the restoration of a viable institutional framework (pp. 584–85). This increase in community feeling, which results from the individual's memory of predisaster institutions and from his perception (clouded by uncertainty) of his current situation relative to the rest of the community (p. 584), suggests to Douty that the remaining stock of "necessity goods" must be distributed in such a way as to allow the victims to survive without leaving the disaster zone (p. 584). Since the prices charged must be within the means of all the victims, and since many victims are without accessible liquid funds, Douty concludes that nonprice rationing schemes must be adopted. Owners of necessity goods are pressured to conform (p. 585).

---

laws, there is no evidence that they typically seek to enforce informal rules of cooperation following a disaster. Some laws (e.g., liability) actually discourage such behavior.

[13] Hirshleifer notes other instances of alliance behavior, such as grueling hours worked by Washington bureaucrats and willingness of soldiers to die. All sorts of people work hard at their jobs, however, and alliance activity would have to be distinguished from more normal drives such as wealth and power. Patriotism, including the willingness to accept a nonzero probability of dying in war, appears to be a useful survival characteristic for groups of individuals and may be explained by alliance activity. On the other hand, it may also be explained by altruism (e.g., protect one's family), self-interest (e.g., earn the benefits of being a hero), duress (e.g., fight or be shot), play, and other motives either singly or in combination. As noted earlier, such things as custom, culture, and the size of the group play the same role under either altruism or alliance.

[14] It may be argued that altruism relates primarily to the redistribution of goods and services whereas alliance concentrates upon the desires or opportunities for continued production. This distinction, suggested to me by Hirshleifer, seems worth pursuing further.

[15] The reader should be reminded that Douty died after he had submitted his paper but before he had an opportunity to respond to referees' comments. M. W. Reder kindly edited the manuscript (see Douty, title footnote, p. 580), but it must be presumed that he hesitated to undertake major changes that might not have been acceptable to the author. Had Douty been able to revise his paper, he might well have avoided some of the criticisms raised below.

[16] Douty indicates that his use of "consensual agreement" corresponds to Hirshleifer's use of "alliance," see fn. 10, p. 585.

According to Douty, "Reconciling this type of behavior with the hypothesis of individual utility maximization . . . presents an interesting theoretical challenge with several different, though not mutually exclusive answers" (p. 586). First, families and friends are linked together by unwritten, mutual insurance contracts against economic adversity. Strangers are included in the coverage as the community, responding to pressure induced by the disaster, acts as an informal reinsurance society to backstop the primary insurance network. Secondly, political, religious, business, and other community leaders recognize that their postdisaster activities are facilitated by such things as stable relationships and a respect for law and order. Accordingly, they press for and obtain community cooperation because they have the ability to reward and to punish. Thirdly, a disaster weakens the enforcement of private property rights.[17] This provides greater scope for community supervision of individual activities, thereby constraining individual choices toward more altruistic behavior.[18] In summary, the main thrust of Douty's argument is that individuals are tied together by an informal network of mutual insurance enforced by social pressure. When a disaster occurs, victims collect.

The informal insurance hypothesis has some initial appeal. Individuals normally sign a variety of formal insurance contracts. In the absence of any evidence to the contrary, it seems reasonable to suppose that they also execute informal insurance agreements. Following a disaster, the payment of compensation would then give rise to seemingly charitable behavior. Although this possibility had already been noted in the literature as a phenomenon additive to charity (see the author, 1968, pp. 528–29), its presentation by Douty as the key explanatory hypothesis is almost convincing on a casual reading. More careful examination, however, uncovers some serious flaws.

Perhaps the most striking flaw is methodological. Several passages in Douty's presentation suggest the inference (for example, p. 587) that the informal insurance hypothesis is superior to the utility interdependence hypothesis on the purely logical ground that it yields the same implications even though it uses fewer initial hypotheses. Indeed, it is difficult to interpret the main thrust of his paper in any other way. The minimum condition necessary to justify such an application of Occam's razor, however, is that both hypotheses yield identical implications. It simply is not enough to argue, as Douty does, that both imply altruism. If some of the lower level implications differ—and, as will be shown later, some do—then the relevance of the competing hypotheses can only be determined empirically. In the absence of such evidence, and none is offered, utility interdependence cannot be ruled out. Moreover, to exclude it would require more rather than fewer initial hypotheses.

Disregarding disaster-related behavior, there is a good deal of evidence that individuals do take others' welfare into consideration. For example, a variety of institutions (for example, most of the agencies receiving support through local United Giver campaigns) annually receive millions of dollars from anonymous contributors who could have given less (say, zero) without decreasing any of their present or future claims to assistance. Similarly, a good deal of insurance (for example, life) is for the benefit of others. This and other behavior is generally attributed to utility interdependence, and the hypothesis is now presented routinely in a variety of texts in economic theory (see, for example, Armen Alchian and William Allen, pp. 166–72; Gary Becker, p. 126).

---

[17] According to Douty, "In the aftermath of a disaster, courts cannot function and soldiers cannot be mobilized" (p. 587). This statement would narrow the analysis to an unnecessarily small category of disasters. In particular, it would exclude most if not all examples of disasters used by Douty. Accordingly, it seems best to disregard it.

[18] It is not clear why this should be the case. As noted earlier, a breakdown of the enforcement mechanism seems to encourage bad behavior. At least in the United States, major disasters typically evoke the mobilization and deployment of National Guard units reportedly to maintain order and prevent looting. This suggests that otherwise bad behavior would have increased to unacceptable levels.

If individual choices normally reflect a concern for the welfare of others, then to exclude utility interdependence in postdisaster periods requires the introduction of additional statements in the theory. Such statements yield a more complex rather than a simpler theory, and their introduction must be justified on methodological and, ultimately, empirical grounds. In particular, it is necessary to show either that the general utility interdependence hypothesis is false or that it is inapplicable in postdisaster periods. Douty does not even question the general applicability of the hypothesis. Indeed, he uses it to explain transfers from donors outside the disaster area (p. 588). Nor does he attempt to show that it is inapplicable in postdisaster periods. He merely offers a substitute hypothesis. But in the absence of any theoretical justification, it is clearly inappropriate—however convenient it might be—to ascribe *any* transfers from without to utility interdependence and *all* transfers from within to insurance or even to informal contracts of all types. Indeed, since information, search, and similar costs are lower for individuals within the disaster area, the latter may be expected to be even more generous—other things being the same—than individuals outside the disaster area.

The viability of any system of contracts depends closely upon the effectiveness of the enforcement mechanism, and Douty's analysis leans entirely upon social pressure for this purpose. There is little indication, however, why a utility-maximizing individual would have the incentive either to meet his own informal obligations or to incur costs to press others to meet their alleged, informal obligations. Under a narrow view of self-interest, an individual will seek to take a free ride. In particular, he will exert pressure on others only if thereby he can increase his receipts as a victim or decrease his contributions as a payer. And in choosing the extent to which he will meet his obligations and the amount and kinds of pressure he will bring to bear on others, he will take into account only *his* gains and *his* costs.

Postdisaster circumstances are ideally suited to free (if not easy) riding. The informal contracts discussed by Douty (and by Hirshleifer) relate groups of individuals (for example, victims and nonvictims) rather than specific individuals. Thus, it would be relatively easy for an individual to claim to have suffered greater losses than he actually had, or to claim to have already contributed, or any combination of these and other strategies intended to maximize his present wealth. No one is responsible for keeping track of who gave what to whom, and no individual (for example, the insured) has the incentive to make sure that another individual (for example, the insurer) is meeting his obligations. Moreover, legal enforcement agencies are constrained by formal laws, particularly those relating to due process and to other safeguards of individual rights. Together with the general disruption of normal information channels and the occupation of enforcement agencies with other tasks (for example, fighting fires, clearing rubble, coordinating outside help), the cost of free riding would be relatively low. Accordingly, it seems reasonable to conjecture that relatively few individuals would be constrained to meet their own obligations or to press others to meet theirs.

Since nonbeneficiaries have the incentive to shirk, social pressure would lack the broad base necessary to make it effective. After all, government is social pressure writ large and endowed with a monopoly on coercion in order to make enforcement viable.[19] Leaders' exhortations are not enough.

Even within the narrow approach chosen by Douty, however, there are no obvious reasons for limiting all informal contracts to insurance, which provides for the compensation of the givers if and only if they in turn become victims. All other informal contracts, such as loans and barter, also have to be taken into account.

Informal loans certainly would account for some seemingly charitable behavior. Immediately following a disaster, lending institutions would find it costlier than usual to establish the credit standing of would-be borrowers. At the higher cost less information is acquired, and both the higher cost and the increased risk are reflected in higher interest rates. Moreover, many borrowers would be seeking relatively small loans for relatively brief periods of time while they sorted out their affairs, and thus formal transactions would be relatively expensive. Under these circumstances, there would be greater scope than normal for negotiating mutually advantageous, informal loans among private individuals.[20] Such loans need not be limited to family, friends, and neighbors, and could profitably be extended to others. Informal loans, particularly if extended in real goods, could easily be mistaken for charity—or for informal insurance.

Informal transfers of ownership also might account for some apparently charitable behavior. For example, the disruption of various market institutions and the resulting increase in transaction costs would encourage some barter in goods and property (for example, blankets for bandages).

Of course, any argument that informal contracts (singly or in combination) fully explain any given transfer of resources must show that the transactions occur at the appropriate market prices. If the prices charged are below their market levels—after taking the appropriate pecuniary and nonpecuniary variables into account except those relating to utility interdependence—then the differential represents a transfer of resources that has to be explained some other way. In such an event, utility interdependence would be a prime candidate to explain the residual.

A more catholic approach to postdisaster altruism, one which consolidates most of the characteristics already noted, might be described as the public finance view of resource redistribution. Following a disaster, informal social arrangements—including some providing for the use of force against unwilling participants—are used to tax certain members of the community and to transfer these resources to others, including victims. One advantage of this approach is that it allows for involuntary transfers.[21] The question of why the informal tax system exists and why its provisions are enforced, however, remains open. The investigator is forced back to theories of the state (see, for example, Anthony Downs), utility interdependence (see, for example, Harold Hochman and James Rodgers; but see also Gordon Tullock), insurance, and so on. Since voluntary transfers apparently account for most postdisaster altruism observed (see Dacy and Kunreuther; Douty), however, this approach will not be pursued further.

---

[19] Liberal disaster bills passed by Congress in recent years do not appear to support the insurance hypothesis. Indeed, the "liberality" of the bills in itself is evidence of the gift component. Such gifts are not inconsistent with the incentive present in democratic (and other) constitutions to use general tax funds to benefit special interest groups.

[20] The availability of low-interest loans and forgiveness grants from various government agencies such as the Small Business Administration would decrease the opportunity for informal loans, but would not eliminate it. Informal loans would still take place to tide individuals over the interim before government and other funds become available, to provide amounts too small to justify formal contracting costs, to finance activities unacceptable to government agencies, and similar options.

[21] Some involuntary transfers (for example, theft and robbery) presumably account for some seemingly charitable behavior. Thus, an individual may be induced to sell goods and services below their open market prices in order to avoid their outright theft or to avoid harm to his person and to his property. Although Douty might disagree (p. 585), it seems best not to force such transactions into the mold of informal insurance.

At this point a few comments regarding business altruism seem in order. It has been recognized for some time (see, for example, Johnson) that under certain market conditions, the traditional owners' wealth-maximization hypothesis implies some altruistic behavior as firms "give away" resources in order to acquire goodwill or avoid badwill. In discussing postdisaster behavior in my 1968 paper, I further conjectured that, under certain ownership arrangements, utility-maximizing decision makers within business firms have the opportunity to satisfy their own taste for the welfare of others by giving away *more* business resources than would otherwise have been the case.

In keeping with his desire to explain postdisaster altruism without involving utility interdependence, Douty asserts that " . . . the postulated taste for altruism is not required if the argument is reframed in terms of the degree of involvement of the firm with the stricken community and the firm's size" (p. 587). To show that altruistic behavior may be deduced from the wealth-maximization hypothesis, however, does not exclude the possibility that at least some business transfers may be charitable.

Douty's analysis, though weak in parts,[22] simply suggests once again that altruistic behavior may be deduced from the standard wealth-maximization hypothesis. In order to rule out the possibility of some charity, however, it is necessary to show either that the implications of the utility maximization hypothesis have been deduced incorrectly or that its antecedent conditions are inapplicable. Ultimately, of course, the validity of the hypothesis can only be questioned empirically. Douty does none of these things.[23]

A final point deserves note. If a firm is to acquire goodwill through altruism, there must be a presumption that charity is a good within the community: the welfare of others matters. "Free" advertising is generally negligible, and narrowly selfish individuals would increase their patronage of "charitable" firms only to the extent that they expected thereby to benefit from *future* disaster-induced transfers. But even if, following a disaster, firms charged lower prices only to their own regular customers (a necessary condition to induce any increase at all in patronage), the present value of the gains expected by prospective customers would be small enough (even if the additional cost of patronizing the donor were zero) that additional patronage would be negligible.

---

[22] For example, the argument that the size and community involvement of the firm matter is vague and largely by analogy. No definition of size is provided, and some firms which fit Douty's description (e.g., brokerage houses with branch offices) do not appear to yield the desired implications.

[23] To buttress his case for the effect of firm size on altruism, Douty notes, fn. 18, p. 587, that the Southern Pacific was the largest benefactor of San Francisco following the earthquake of 1906 and that the Safeway stores in Alaska did not increase the prices of staples (some prices were actually lowered) following the earthquake of 1964. These observations, however, are too casual to be meaningful. For example, was the Southern Pacific the "largest" firm in San Francisco? Was the Safeway's rank order as benefactor similar to its rank order for size? How did other firms behave? Were other factors actually at work? Thus, Johnson's tentative study of corporate philanthropy does not support Douty's conjecture regarding size. His results indicate that the ratio of charitable contributions to profits does not vary directly with asset size, and suggest that other variables (say, degree of competition) account for the behavior he observed.

IV. Range of Application of the Competing Hypotheses

The analysis so far suggests that alliance activities, informal contracts, and utility interdependency may all account for some postdisaster altruism. Obviously, it would be helpful to establish the range of application of these hypotheses. Although a definitive answer must wait until more rigorous empirical evidence becomes available, various theoretical and empirical considerations encourage a preliminary answer.[24]

Utility interdependence already appears to provide the major if not the sole explanation for voluntary resource transfers from individuals *outside* the disaster area. The case for informal insurance is particularly weakened by enforcement problems and by the lack of incentive to make anonymous contributions, and even Douty acknowledges that pure altruism accounts for much aid from outsiders (p. 588). Alliance activity similarly would be strained to explain all such behavior (for example, foster parent programs).

Turning to altruistic behavior *within* the disaster area, a crucial characteristic which distinguishes insurance from charity is that the individuals who participate in an insurance scheme expect to be compensated should they become victims later on. Among other things, this implies that an individual will reveal a strong preference to help those victims who offer the best evidence of being able to compensate him in the future should the occasion arise. A donor, on the other hand, will prefer to aid those who, *ceteris paribus*, seem to have suffered most from the disaster. As a result, an insurer will prefer to aid young, healthy, wealthy individuals who are temporarily distressed rather than old, ill, and poor individuals: the latter are much less likely to be around and to be able to compensate the giver if and when the occasion arises. Although even the old, the weak, and the poor eventually might be aided under some sort of an informal assigned risk plan enforced by social pressure (how?), the expectation is that they would be helped last and least. A donor, on the other hand, will reveal a stronger preference for these people—controlling for family, friendship, and similar ties—exactly because they are less able to take care of themselves. Theoretical considerations noted earlier suggest that the insurance hypothesis is not likely to prevail. This conclusion is supported by strong evidence from psychological studies indicating that the single most important characteristic of recipients is dependency (see Krebs, pp. 277–79).

Alliance activity, particularly of the kind induced by custom, appears to be indistinguishable from charity. Indeed, as noted earlier, it could be used to explain it. Altogether, utility interdependence appears to have significant explanatory power within the disaster area as well.

Other informal contracts, where the parties to the exchange belong to amorphous groups and fulfillment of the contract is essentially random, are subject to the same limitations of the informal insurance program. Transfers of a more traditional nature, where the terms of the contract are specified and the individual parties to the transaction are clearly identified but simply do not bother to sign a formal contract, would give rise to behavior which to an uninformed bystander might seem to be charitable. It should be possible, however, to identify such events and to treat them accordingly.

---

[24] Douty's position on the range of application of the informal insurance hypothesis at best is equivocal. Although the whole thrust of his argument is that it explains the behavior observed without reference to utility interdependence, he concludes the theoretical presentation of his case with the statement that "Nothing further need be said about the characteristics of anyone's utility function, though, 'true altruism' may be present in postdisaster situations" (p. 587). But true altruism, as defined by Douty, arises from utility interdependence. Accordingly, something further *is* being said about individual utility functions: they are interdependent.

Additional research, particularly at the empirical level, obviously is desirable to clarify matters further. Beyond some point, however, the explanation of altruism properly belongs to psychology and sociology. Unfortunately, the evidence here at present is also inadequate. Psychological studies have begun to identify a number of significant characteristics of givers and of recipients,[25] but researchers have not yet found a way to measure the motivational base of altruism, and the concept of altruism itself is still unclear (see Krebs, p. 297). Sociological explanations of altruism have focused upon norms of social responsibility (benefactors should help those who need help) and of reciprocity (beneficiaries should help, or not harm, those who have helped them) with mixed empirical support (see Krebs, pp. 294–97). Interestingly enough for present purposes, proponents of the reciprocity hypothesis (a norm which characterizes at least Douty's position) note only one exception to its range of application: emergency situations, where the norm of giving holds (see Krebs, pp. 294–95). Why such norms should exist—if, indeed, they do—awaits further research.[26]

V. Conclusions

Economic explanations of postdisaster cooperation initially focused upon utility interdependence and business wealth-maximization. Hirshleifer, however, suggested that these explanations failed to account for some of the behavior observed and proposed alliance activity to fill the gap. Douty sought to rule out utility interdependence entirely and to explain all postdisaster cooperation in terms of narrowly conceived utility maximizing behavior.

The papers by Douty and by Hirshleifer suggest that the consequences of various informal contracts deserve to be taken into account. Informal insurance and alliance activity, however, do not appear to have sufficient power to explain all the behavior observed. The presence of utility interdependence under normal conditions, the lack of an effective mechanism to enforce informal contracts, and evidence of disaster-connected charity (for example, anonymous gifts from outside the community) suggest that utility interdependence matters in postdisaster situations.

Different instances of postdisaster cooperation may have different causes, and it would be difficult to defend any one single explanation. Presumably, some transfers may be attributed to utility interdependence, others to informal contracts of various sorts (including loans and transfers of ownership shares), and still others, perhaps antecedent to utility interdependence, to such things as alliance activity.[27]

---

[25] Krebs' recent review of the psychological and related sociological literature on altruism reports the following main conclusions. The results of experiments in which situational variables were manipulated indicate that individuals become more altruistic after experiencing success and competence as well as guilt and shame and after observing altruistic acts by others (pp. 264–77). In particular, there is evidence that individuals who perceive themselves to be competent in emergencies give more help in postdisaster situations (pp. 264, 266). The most important characteristic of a recipient in eliciting aid is dependency (pp. 277–79); interpersonal attractiveness is the only other significant variable (pp. 279–81). Although personality traits of both givers and recipients do not seem to matter (pp. 281–86), several social and demographic variables do: sex, age, ordinal position, social class, and nationality seem to affect altruism (pp. 286–94). Regarding the social roles of recipients, it appears that individuals give to those who are similar to themselves, are prestigious, and are more likely to benefit them in the future. Moreover, the receipt of favors elicits a general disposition toward altruism described as generalized reciprocity (pp. 295–97).

[26] According to Krebs, "Little psychological research has examined the norm of giving, even though it may have been antecedent to many altruistic situations" (p. 295).

[27] The reason why any good enters any individual's utility function ultimately must be selfish. The goal(s) which the good satisfies are the individual's goals, however determined, and the individual chooses the extent to which alternative goals are to be fulfilled, subject to the constraints which limit his choices.

Nevertheless, although utility interdependence may not be sufficient to explain all postdisaster cooperation, it certainly appears to be necessary—a statement which does not appear to hold for the other alternatives considered.

REFERENCES

A. A. Alchian and W. R. Allen, *Exchange and Production: Theory in Use,* 2d ed., Belmont 1969.

G. S. Becker, *Economic Theory,* New York 1971.

D. C. Dacy and H. R. Kunreuther, *The Economics of Natural Disasters: Implications for Federal Policy,* New York 1969.

L. De Alessi, "A Utility Analysis of Postdisaster Cooperation," *Papers in Non-Market Decision-Making* (now *Public Choice*), Fall 1967, *3,* 85–90.

———, "The Utility of Disasters," *Kyklos,* 1968, *21,* 525–32.

C. M. Douty, "Disasters and Charity: Some Aspects of Cooperative Economic Behavior," *Amer. Econ. Rev.,* Sept. 1972, *62,* 580–90.

A. Downs, *An Economic Theory of Democracy,* New York 1957.

J. Hirshleifer, "Disaster Behavior: Altruism or Alliance?" unpublished paper, 1967.

H. M. Hochman and J. D. Rodgers, "Pareto Optimal Redistribution," *Amer. Econ. Rev.,* Sept. 1969, *59,* 542–57.

O. Johnson, "Corporate Philanthropy: An Analysis of Corporate Contributions," *J. Bus., Univ. Chicago,* Oct. 1966, *39,* 489–504.

D. L. Krebs, "Altruism—An Examination of the Concept and a Review of the Literature," *Psychological Bull.,* Apr. 1970, *73,* 258–302.

K. Lorenz, *King Solomon's Ring,* New York 1952.

M. Olson, Jr. and R. Zeckhauser, *An Economic Theory of Alliances,* RAND Corporation, RM-4297-ISA, Santa Monica, Oct. 1966.

G. Sjoberg, "Disasters and Social Change," in G. W. Barker and D. W. Chapman, eds., *Man and Society in Disaster,* New York 1962.

G. Tullock, "The Charity of the Uncharitable," *Western Econ. J.,* Dec. 1971, *9,* 379–92.

Polluters' Profits and Political Response: Direct Controls Versus Taxes

By James M. Buchanan and Gordon Tullock*

Economists of divergent political persuasions agree on the superior efficacy of penalty taxes as instruments for controlling significant external diseconomies which involve the interaction of many parties. However, political leaders and bureaucratic administrators, charged with doing something about these problems, appear to favor direct controls. Our purpose in this paper is to present a positive theory of externality control that explains the observed frequency of direct regulation as opposed to penalty taxes or charges. In the public-choice theory of policy,[1] the interests of those who are subjected to the control instruments must be taken into account as well as the interests of those affected by the external diseconomies. As we develop this theory of policy, we shall also emphasize an elementary efficiency basis for preferring taxes and charges which heretofore has been neglected by economists.

I

Consider a competitive industry in long-run equilibrium, one that is composed of a large number of n identical producing firms. There are no productive inputs specific to this industry, which itself is sufficiently small relative to the economy to insure that the long-run supply curve is horizontal. Expansions and contractions in demand for the product invoke changes in the number of firms, each one of which returns to the same least-cost position after adjustment. Assume that, from this initial position, knowledge is discovered which indicates that the industry's product creates an undesirable environmental side effect. This external diseconomy is directly related to output, and we assume there is no technology available that will allow alternative means of producing the private good without the accompanying public bad. We further assume that the external damage function is linear with respect to industry output; the same quantity of public bad per unit of private good is generated regardless of quantity.[2] We assume that this damage can be measured and monitored with accuracy.

This setting has been deliberately idealized for the application of a penalty tax or surcharge. By assessing a tax (which can be computed with accuracy) per unit of output on all firms in the industry, the government can insure that profit-maximizing decisions lead to a new and lower industry output that is Pareto optimal. In the short run, firms will undergo losses. In the long run, firms will leave the industry and a new equilibrium will be reached when remaining firms are again making normal returns on investment. The price of the product to consumers will have gone up by the full amount of the penalty tax.

---

* Center for Study of Public Choice, Virginia Polytechnic Institute and State University. We wish to thank the National Science Foundation for research support. Needless to say, the opinions expressed are our own.

[1] Charles Goetz imposes a public-choice framework on externality control, but his analysis is limited to the determination of quantity under the penalty-tax alternative.

[2] This assumption simplifies the means of imposing a corrective tax. For some of the complexities, see Otto Davis and Andrew Whinston and Stanislaw Wellisz.

No one could dispute the efficacy of the tax in attaining the efficient solution, but we should note that in this setting, the same result would seem to be equally well insured by direct regulation. Policy makers with knowledge of individual demand functions, the production functions for firms and for the industry, and external damage functions, could readily compute and specify the Pareto-efficient quantity of industry output.[3] Since all firms are identical in the extreme model considered here, the policy makers could simply assign to each firm a determinate share in the targeted industry output. This would then require that each firm reduce its own rate of output by X percent, that indicated by the difference between its initial equilibrium output and that output which is allocated under the socially efficient industry regulation.[4]

Few of the standard arguments for the penalty tax apply in this setting. These arguments have been concentrated on the difficulties in defining an efficient industry output in addition to measuring external damages and on the difficulty in securing data about firm and industry production and cost functions. With accurately measured damage, an appropriate tax will insure an efficient solution without requiring that this solution itself be independently computed. Or, under a target or standards approach, a total quantity may be computed, and a tax may be chosen as the device to achieve this in the absence of knowledge about the production functions of firms.[5]

In the full information model, none of these arguments is applicable. There is, however, an important economic basis for favoring the penalty tax over the direct control instrument, one that has been neglected by economists. The penalty tax remains the preferred instrument on strict efficiency grounds, but, perhaps more significantly, it will also facilitate the enforcement of results once they are computed.[6] Under the appropriately chosen penalty tax, firms attain equilibrium only at the efficient quantity of industry output. Each firm that remains in the industry after the imposition of the tax attains long-run adjustment at the lowest point on its average cost curve only after a sufficient number of firms have left the industry. At this equilibrium, there is no incentive for any firm to modify its rate of output in the short run by varying the rate of use of plant or to vary output in the long run by changing firm size. There is no incentive for resources to enter or to exit from the industry. So long as the tax is collected, there is relatively little policing required.

This orthodox price theory paradigm enables the differences between the penalty-tax instrument and direct regulation to be seen clearly. Suppose that, instead of levying the ideal penalty tax, the fully informed policy makers choose to direct all firms in the initial competitive equilibrium to reduce output to the assigned levels required to attain the targeted efficiency goal for the industry. No tax is levied. Consider Figure 1, which depicts the situation for the individual firm. The initial competitive equilibrium is attained when each firm produces an output, q_i. Under regulation it is directed to produce only q_0, but no tax is levied. At output q_0, with an unchanged number of firms, price is above

---

[3] See Allen Kneese and Blair Bower, p. 135.

[4] No problems are created by dropping the assumption that firms are identical so long as we retain the assumption that production functions are known to the regulator.

[5] This is the approach taken by William Baumol, who proposes that a target level of output be selected and a tax used to insure the attainment of this target in an efficient manner.

[6] See George Hay. His discussion of the comparison of import quotas and tariffs on oil raises several issues that are closely related to those treated in this paper.

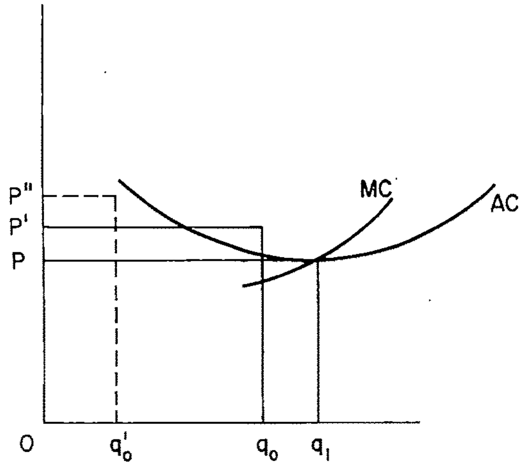


FIGURE 1

marginal cost (for example price is at P'). Therefore, the firm is not in short-run equilibrium, and would if it could expand output within the confines of its existing plant. More importantly, although each firm will be producing the output quota assigned to it at a somewhat higher cost than required for efficiency reasons, there may still be an incentive for resources to enter the industry. The administrator faces a policing task that is dimensionally different from that under the tax. He must insure that individual firms do not violate the quotas assigned, and he must somehow prevent new entrants. To the extent that the administrator fails in either of these tasks, the results aimed for will not be obtained. Output quotas will be exceeded, and the targeted level of industry production overreached.

If the administrator assigns enforceable quotas to existing firms and successfully prevents entrants, the targeted industry results may be attained, but there may remain efficiency loss since the industry output will be produced at higher average cost than necessary if firms face U-shaped long-run average cost curves. Ideally, regulation may have to be accompanied by the assignment of full production quotas to a selected number of the initial firms in the industry. This policy will keep these favored firms in marginal adjustment with no incentives for in-firm adjustments that might defeat the purpose of the regulation. But even more than under general quota assignment there will be strong incentives for firms to enter the industry and to secure at least some share of the rents that the restriction of industry output generates. If the response to this pressure should be that of reassigning quota shares within the unchanging and targeted industry output so as to allow all potential entrants some share, while keeping all firms, actual and potential, on an equal quota basis, the final result may be equivalent to the familiar cartel equilibrium. No firm will be earning more than normal returns, but the industry will be characterized by too many firms, each of which produces its assigned output inefficiently.

II

When we examine the behavioral adjustments to the policy instruments in the manner sketched out above, a theory of policy emerges. Regulation is less desirable on efficiency grounds even in the presence of full information, but this instrument will be preferred by those whose behavior is to be subjected to either one or the other of the two policy instruments. Consider the position of the single firm in the fully competitive industry, depicted in Figure 1. Under the imposition of the tax, short-run losses are necessarily incurred, and the firm reattains normal returns only after a sufficient number of its competitors have shifted resources to other industries. The tax reduces the present value of the firm's potential earnings stream, whether the particular firm remains in the industry after adjustment or withdraws its investment and shifts to alternative employment. In terms of their own private interests, owners of firms in the industry along with employees will oppose the tax. By contrast, under regulation firms may well secure pecuniary gains from the imposition of direct controls that reduce total industry output. To the extent that the restriction is achieved by the assignment of production quotas to existing firms, net profits may be present even for the short term and are more likely to arise after adjustments in plant. In effect, regulation in this sense is the directional equivalent of cartel formation provided that the individual firm's assigned quota falls within the limited range over which average cost falls below price. Such a range must, of course, exist, but regulatory constraints may possibly be severe enough to shift firms into positions where short-term, and even possibly long-term, losses are present, despite increased output price. Such a result is depicted by a restriction to q_0' in Figure 1, with price at P''.

Despite the motivation which each firm has to violate assigned quotas under regulation, it remains in the interest of firms to seek regulatory policy that will enforce the quotas. If existing firms foresee the difficulty of restricting entry, and if they predict that governmental policy makers will be required to accommodate all entrants, the incentive to support restriction by regulation remains even if its force is somewhat lower. In final cartel equilibrium, all the firms will be making no more than normal returns. But during the adjustment to this equilibrium, above-normal returns may well be available to all firms that hold production quotas. Even if severe restriction forces short-term losses on firms, these losses will be less than those under the tax. Rents over this period may well be positive, and even if negative, they will be less negative than those suffered under the tax alternative. Therefore, producing firms will always oppose any imposition of a penalty tax. However, they may well favor direct regulation restricting industry output, even if no consideration at all is given to the imposition of a tax. And, when faced with an either/or choice, they will always prefer regulation to the tax.

III

There is a difference between the two idealized solutions that has not yet been discussed, and when this is recognized, the basis of a positive hypothesis about policy choice may appear to vanish. Allocationally, direct regulation can produce results equivalent to the penalty tax, providing that we neglect enforcement cost differentials. *Distributionally*, however, the results differ. The imposition of tax means that government collects revenues (save in the case where tax rates are prohibitive) and these must be spent. Those who anticipate benefits from the utilization of tax revenues, whether from the provision of publicly supplied goods or from the reduction in other tax levies, should prefer the tax alternative and they should make this preference known in the political process. To the extent that the beneficiaries include all or substantially all members of the community, the penalty tax should carry the day. Politicians, in responding to citizenry pressures, should heed the larger number of beneficiaries and not the disgruntled members of one particular industry. This political choice setting is, however, the familiar one in which a small, concentrated, identifiable, and intensely interested pressure group may exert more influence on political choice making than the much larger majority of persons, each of whom might expect to secure benefits in the second order of smalls.

There is an additional reason for predicting this result with respect to an innovatory policy of externality control. The penalty tax amounts to a legislated change in property rights, and as such it will be viewed as confiscatory by owners and employees in the affected industry. Legislative bodies, even if they operate formally on majoritarian principles, may be reluctant to impose what seems to be punitive taxation. When, therefore, the regulation alternative to the penalty tax is known to exist, and when representatives of the affected industry are observed strongly to prefer this alternative, the temptation placed on the legislator to choose the direct control policy may be overwhelming, even if he is an economic theorist and a good one. Widely accepted ethical norms may support this stance; imposed destruction of property values may suggest the justice of compensation.[7]

If policy alternatives should be conceived in a genuine Wicksellian framework, the political economist might still expect that the superior penalty tax should command support. If the economist ties his recommendation for the penalty tax to an accompanying return of tax revenues to those in the industry who suffer potential capital losses, he might be more successful than he has been in proposing unilateral or one-sided application of policy norms. If revenues are used to subsidize those in the industry subjected to capital losses from the tax, and if these subsidies are unrelated to rates of output, a two-sided tax subsidy arrangement can remove the industry source of opposition while still insuring efficient results. In this respect, however, economists themselves have failed to pass muster. Relatively few modern economists who have engaged in policy advocacy have been willing to accept the Wicksellian methodological framework which does, of course, require that some putative legitimacy be assigned to rights existent in the status quo.[8]

IV

To this point we have developed a theory of policy for product-generated external diseconomies, the setting which potentially counterposes the interest of members of a single producing industry against substantially all persons in the community. External diseconomies may, however, arise in consumption rather than in production, and these may be general. For purposes of analysis, we may assume that all persons find themselves in a situation of reciprocal external diseconomies. Traffic congestion may be a familiar case in point.

The question is one of determining whether or not persons in this sort of interaction, acting through the political processes of the community, will impose on *themselves* either a penalty tax or direct regulation. We retain the full information assumption introduced in the production externality model. For simplicity here, consider a two-person model in which each person consumes the same quantity of good or carries out the same quantity of activity in the precontrol equilibrium, but in which demand elasticities differ. Figure 2 depicts the initial equilibrium at E with each person consuming quantity Q. The

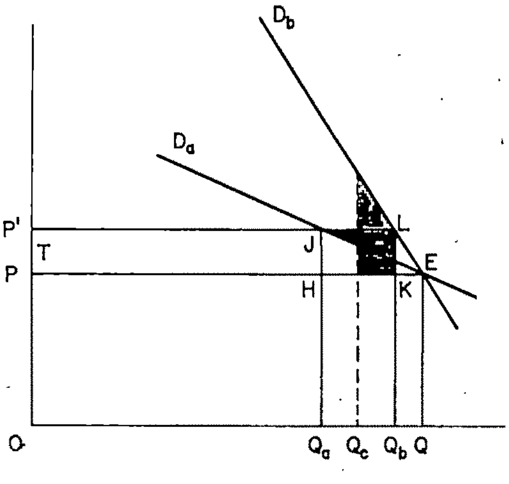


FIGURE 2

---

[7] For a comprehensive discussion of just compensation, see Frank Michelman.

[8] For a specific discussion of the Wicksellian approach, see Buchanan (1959).

existence of the reciprocal external diseconomy is discovered. The community may impose an accurately measured penalty tax in the amount T, in which case A will reduce consumption to Q_a and B will reduce consumption to Q_b. Total consumption is reduced from $2Q$ to (Q_a+Q_b), but both A and B remain in equilibrium. At the new price P', which includes tax, neither person desires to consume more or less than the indicated quantities. The government collects tax revenues in the amount $[2(PP'JH)+HJLK]$. Alternatively, the community may simply assign a restricted quantity quota to each person. If the government possesses full information about demand functions it can reduce A's quota to Q_a, and B's quota to Q_b, securing results that are allocatively identical to those secured by the tax. However, under the quota, both A and B will find themselves out of equilibrium; both will, if allowed quantity adjustment, prefer to expand their rate of consumption.

It will be useful to examine the ideal tax against the quota scheme outlined above, which we may call the idealized quota scheme. If individuals expect no returns at all from tax revenues in the form of cash subsidies, public goods benefits, or reductions in other taxes, both A and B will clearly prefer the direct regulation. The loss in consumers' surplus under this alternative is small relative to that which would be lost under the penalty tax. Each person willingly trades off marginal quantity adjustment for the more favorable inframarginal terms offered under direct regulation, given our assumptions that both instruments achieve the same overall externality control objective.

Under extreme fiscal illusion, individuals may ignore benefits from tax revenues, but consistent methodological precept requires that we allow persons to recognize the benefit side of the fiscal account, at least to some degree. Let us allow all revenues under the penalty tax to be returned in equal shares to all taxpayers. Under this arrangement, each person expects to get back one-half of the amount measured as indicated above for Figure 2. Simplifying, each expects to get back the amount $PP'JH$, which he personally pays in, plus one-half of the amount measured by the rectangle $JHKL$, all of which is paid in by B. From an examination of Figure 2, it is clear that individual A will favor the penalty tax under these assumptions. The situation for individual B is different; he will prefer direct regulation. He will secure a differential gain measured by the horizontally shaded area in Figure 2, which is equal to the differential loss that individual A will suffer under this alternative. The policy result, insofar as it is influenced by the two parties, is a standoff under this idealized tax and idealized quota system comparison.

For constitutional and other reasons, control institutions operating within a democratic order could scarcely embody disproportionate quota assignments. A more plausible regulation alternative would assign quotas proportionate to initial rates of consumption, designed to reduce overall consumption to the level indicated by target criteria. The comparison of this alternative with the ideal tax arrangement is facilitated by the construction of Figure 2 where the initial rates of consumption are equal. In this new scheme, each person is assigned a quota Q_c, which he is allowed to purchase at the initial price P. We want to compare this arrangement with the ideal tax, again under the assumption that revenues are fully returned in equal per head subsidies. As in the first scheme, both persons are in disequilibrium at quantity Q_c and price P. The difference between this model and the idealized quota scheme lies in the fact that at Q_c, the marginal evaluations differ as between the two persons. There are unexploited gains from trade, even under the determined overall quantity restriction.

It will be mutually advantageous for the two persons to exchange quotas and money, but, at this point, we assume that such exchanges do not take place, either because they are prohibited or because transactions costs are too high. Individual *A* will continue to favor the tax alternative but his differential gains will be smaller than under the idealized quota scheme. In the model now considered, *A*'s differential gains under the ideal tax are measured by the blacked-in triangle in Figure 2. Individual *B* may or may not favor the quota, as in the earlier model. His choice as between the two alternatives, the ideal tax on the one hand and the restriction to Q_c at price *P* on the other, will depend on the comparative sizes of the two areas shown as horizontally and vertically shaded in Figure 2. As drawn, he will tend to favor the quota scheme, but it is clearly possible that the triangular area could exceed the rectangular one if *B*'s demand curve is sufficiently steep in slope. In any case, the choice alternatives for both persons are less different in the net than those represented by the ideal tax and the idealized quota.

While holding all of the remaining assumptions of the model, we now drop the assumption that no exchange of quotas takes place between *A* and *B*. To facilitate the geometrical illustration, Figure 3 essentially blows up the relevant part of Figure 2. With each party initially assigned a consumption quota of Q_c, individual *A* will be willing to sell units to individual *B* for any price above his marginal evaluation. Hence, the lowest possible supply price schedule that individual *B* confronts is that shown by the line *RL* in Figure 3. The maximum price that individual *B* is willing to pay for additional units of quota is his marginal evaluation, shown by *SL*. The gains-from-trade are measured by the triangular area *RLS*. The distribution of

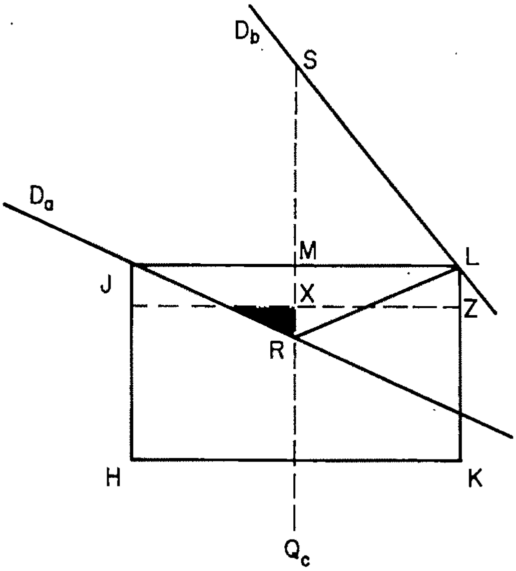


FIGURE 3

these gains will, of course, be settled in the strict two-man setting by relative bargaining skills, but let us assume that individual *B*, the buyer, wants to purchase consumption quota units from *A*, but also to do so in such a way that individual *A* will come to prefer this system over the tax. To accomplish this, he must insure that *A* gets a share of the net gains at least equal to the area *RML* on Figure 3. Individual *B*, the buyer, retains gains of *MSL* under this division of the spoils. But in this arrangement, both persons are indifferent as between the policy alternatives. The system is on the Pareto frontier, and the quota scheme plus the exchange process produces allocative and distributive results identical to those generated under the ideal tax. This becomes the analogue of the Coase theorem in the context that we are examining.[9]

V

These somewhat inconclusive results may seem to provide anything but a positive theory of policy akin to that presented with respect to production externalities. The comparisons are, however, a necessary stage in developing such a theory. Recall that we have made these comparisons under the most favorable possible assumption concerning anticipated return of revenues under the penalty tax. In the real world, individuals will not anticipate that these will be returned dollar-for-dollar, and they will tend to place at least some discount on the value of benefits that they expect.

---

[9] See Ronald Coase. For a related extension of the Coase theorem, see Buchanan (1973).

Let us say that each person expects an aggregate benefit value of only 80 cents on the dollar from tax revenues collected under the penalty tax. Consider what this single change does to the results of the last comparison made, that which involves proportionate quota assignments along with a free market in quotas. In this case, individual B, the buyer, can offer individual A, the seller, more than the amount required to make him prefer the quota alternative, while himself continuing to secure differential benefit under this alternative. Individual A's differential gains from the ideal penalty tax are reduced to the shaded area in Figure 3. By paying individual A the amount measured by RML, he has improved A's position relative to the penalty tax. And, in the process, he has retained for himself a differential gain measured by the area $MXZL$. Both persons in full knowledge of the alternatives will prefer the quota system, and political leaders will presumably respond by opting for regulation.

The same reasoning can readily be extended to apply to any quota system. In the idealized quota assignment first considered, we demonstrated that one person would favor the penalty tax and the other the quota. Individual A, who favors the penalty tax, loses no consumer's surplus, and he does expect to secure an income transfer through the return of tax revenues. When we modify the assumptions concerning expectations of the value of returned revenues or benefits, however, this conclusion need not hold. Individual A will, of course, expect to get back in benefits some part of the tax revenues paid in by B that is in excess of that contributed by A himself. If, however, individual A applies the same discount factor to all revenues collected, the deadweight loss may more than offset the income transfer effect. Examination of Figure 2 indicates that under the 80 percent assumption, one-fifth of the area measured by $PP'JH$ will represent deadweight loss to A from the revenues that he pays in. This deadweight loss may well be larger than the measure of the income transfer that he expects, which amounts to 80 percent of the horizontally shaded area in Figure 2. Once we introduce any plausible discount factor into the expectation of individuals concerning the return of tax revenues, it is relatively easy to demonstrate situations under which both persons may be led by private self-interest to favor the direct regulation alternative.

VI

We have developed a positive theory of externality control policy for both the production and consumption interactions under highly abstract and simplified models which allow us to isolate influences on policy formation which have been neglected. Decisions on the alternative policy instruments in democratic governments are surely influenced by the preferences of those who are subjected to them. The public-choice approach, which concentrates attention on the individual's choice as between policy instruments, allows us to construct hypotheses that explain the prevalence of direct regulation.[10] For economists who continue to support the penalty tax alternative, the analysis suggests that they had best become good Wicksellians and begin to search out and invent institutional arrangements that will make the penalty tax acceptable to those who are primarily affected.

---

[10] Much of the analysis developed in this paper can be applied more or less directly to policy alternatives proposed in the energy crisis of late 1973 and early 1974. For such application, see Buchanan and Nicolaus Tideman.

REFERENCES

W. J. Baumol, "On Taxation and the Control of Externalities," *Amer. Econ. Rev.*, June 1972, *62*, 307–22.

J. M. Buchanan, "Positive Economics, Welfare Economics, and Political Economy," *J. Law Econ.*, Oct. 1959, *2*, 124–38.

———, "The Coase Theorem and the Theory of the State," *Natur. Resources J.*, Oct. 1973, *13*, 579–94.

——— and **N. Tideman**, "Gasoline Rationing and Market Pricing: Public Choice in Political Democracy," research pap. no. 808231-1-12, Center for Study of Public Choice, Virginia Polytechnic Inst. and State Univ., Jan. 1974.

R. H. Coase, "The Problem of Social Cost," *J. Law Econ.*, Oct. 1960, *3*, 1–44.

O. A. Davis and A. Whinston, "Externalities, Welfare, and the Theory of Games," *J. Polit. Econ.*, June 1962, *70*, 241–62.

C. J. Goetz, "Political Equilibrium vs. Economic Efficiency in Effluent Pricing," in J. R. Conner and E. Loehman, eds., *Economic Decisionmaking for Environmental Control*, Gainesville 1973.

G. A. Hay, "Import Controls on Foreign Oil: Tariff or Quota?," *Amer. Econ. Rev.*, Sept. 1971, *61*, 688–91.

A. V. Kneese and B. T. Bower, *Managing Water Quality: Economics, Technology, Institutions*, Baltimore 1968.

F. J. Michelman, "Property Utility, and Fairness: Comments on the Ethical Foundations of 'Just Compensation' Law," *Harvard Law Rev.*, Apr. 1967, *80*, 1165–1257.

S. Wellisz, "On External Diseconomies and the Government-Assisted Invisible Hand," *Economica*, Nov. 1964, *31*, 345–62.

Durability of Consumption Goods Reconsidered

By TEDDY T. SU*

Recent progress in the durability analysis of consumer durables has been inspired by the theory of investment.[1] In 1961 Henry Miller, Jr. applied the theory of investment to a study of optimal allocation of consumption expenditures between durables and nondurables, and proposed an optimum replacement policy for consumer durables. Subsequently, David Martin discussed the effects on durability of monopolization of a perfectly competitive market. More recently, E. Kleiman and T. Ophir, David Levhari and T. N. Srinivasan, Richard Schmalensee (1970), and Peter Swan successively examined the durability of consumer durables and compared the optimum choice of durability under perfect competition and pure monopoly. Contrary to the conclusions of Levhari and Srinivasan, Swan concluded in his recent papers that a monopolist would produce goods of the same durability as competitive firms. It should be noted that Swan made his analysis using parts of the theoretical apparatus developed by Friedrich and Vera Lutz.[2] Moreover, Swan's ingenious idea of transforming a profit-maximization problem into a cost-minimization problem seems to coincide with Vernon Smith's early argument.[3] The only flaw in Swan's analysis is that the durabil-

---

* Private economic consultant, Washington, D.C. Much of the work on this paper was completed at the University of South Carolina. I wish to express my thanks to Robert Carlsson, Elwood Shomo, an anonymous referee, and the managing editor for useful suggestions.

[1] The first mathematical model on durability analysis of durable equipment is found in Wicksell's discussion of "Akerman's Problem," which was published in 1923 and translated into English in 1934. Later on, durability analysis became an important part of investment theory as seen in the works of Friedrich and Vera Lutz, and Vernon Smith. Although consumer durables and capital goods have some characteristics in common, discussion of the durability or replacement problem seems to refer solely to capital goods prior to 1960.

[2] The Lutzes discussed the nature of economic horizon and took profit maximization as a firm's objective, and so did Swan. One of the objective functions used by the Lutzes, p. 120, is

$$G = \left[\int_0^n Re^{-rt}dt - \phi(n)\right](1 + e^{-rn} + e^{-2rn} + \dots)$$
$$= \frac{R}{r} - \frac{\phi(n)}{1 - e^{-rn}}$$

Where G = goodwill or net profit, R = price of the product or service, n = lifetime of the durable, $\phi(n)$ = cost of the durable or service, and r = interest rate. "So long as the rate of interest is a constant, the condition for the optimum lifetime of each durable good in the infinite chain runs entirely in terms of costs" (Lutz and Lutz, p. 121). The condition for the optimum durability of durables in an infinite horizon provided by the Lutzes, p. 120, is

$$\phi'(n) = \frac{r}{(1 - e^{-rn})}\left[\phi(n)e^{-rn}\right]$$

Except for notation, Swan when dealing with the optimum durability of the one-hoss shay used the same objective function and arrived at the same condition. See Swan (1970), p. 886.

[3] Criticizing the profit maximization approach taken by the Lutzes, Smith (1959) argued, "By definition . . . the act of replacement should not affect any output dimension . . . replacement is concerned with minimizing the cost of producing a given output" (p. 73). Thus Smith regards optimum replacement as least cost production. Smith's model so formulated has aimed at minimizing the annual cost of the service flow in an infinite horizon. He states: "This earlier theoretical literature of replacement theory is almost wholly concerned with the replacement of machines with units of identical type. It is therefore puzzling to find such tenacious adherence to a profit maximization formulation of the model, since, in the absence of technological changes in equipment, the replacement decision cannot possibly affect either price or output" (1961, p. 131). Swan appears to be unaware of Smith's argument when he endeavors to prove that the firm's sales revenue is independent of durability choice.

ity problem has been examined merely from the supply side, i.e., the standpoint of producers. A similar flaw is found in the durability analysis of the other writers mentioned above. Although the consumer's version or demand side of durability analysis was discussed briefly by Hans Brems in 1951 and revised in his more recent book, it remains to be generalized and integrated with the supply side. The objective of this paper is to reconsider the supply side of durability analysis in conjunction with the demand side neglected by earlier writers. A similar objective seems to be embodied in papers by Schmalensee (1974) and Richard Parks which incorporate the consumer's decision in the formulation of optimum durability. Parks considers transaction and operating costs in connection with the purchase and use of one-hoss shay durables while Schmalensee introduces the optimum choice of maintenance effort in a model dealing with "radioactive decay" durables, provided that a market is composed of homogeneous consumers. Under the same essential assumptions,[4] Swan's argument, while supported by Parks, is rebutted by Schmalensee (1974), who has proposed that durability is dependent on market structure in the sense that monopoly rental policies, while yielding the same result as competitive sales and rental policies, differ from monopoly sales policies. Schmalensee's propositions are supported by the results of the present analysis centering on the consumer's choice of durability. However, this study shows that the monopolist's choice of durability could be the same under the sales policy as under the rental policy though it would distort consumers' choice.

In light of the Lutzes' analysis, consumer durables are classified into the one-hoss shay and diminishing-efficiency varieties. Since the lifetime of one-hoss shay durables is by definition a technical parameter and cannot possibly be adjusted by consumers, only diminishing-efficiency durables will be discussed in this paper.

I. The Demand Side of the Durability Problem

If firms rent rather than sell durables, the demand side of durability analysis will be irrelevant, because under the rental policy firms determine both the quality and the optimum time for replacement of durables. Hence, this section will deal only with those durables which are sold by firms to consumers. The quality of a durable is determined by the firms, but buyers of the durable may choose the optimum time of replacement.[5] Note that the lifetime of a given durable may be extended by repairs. Let us assume that consumers have purchased a durable good of quality α. The consumer's objective is to minimize the per unit expense of service rendered by the durable. The price of the durable, P, is assumed to be positive and to be a monotonically increasing function of α.[6] The variable α is an ordinal quality index. In mathematical notation,

---

[4] The present study retains the following assumptions made by Levhari and Srinivasan and adopted by Schmalensee, Swan, and Parks: (i) An economy is in the stationary state of long-run equilibrium; (ii) Returns to scale are constant; (iii) Cost functions of firms are identical; and (iv) Both consumers and producers have perfect foresight. Recently Sieper and Swan have relaxed assumption (i), and Parks has apparently removed assumption (iv) when he considers heterogeneous consumers and monopoly discrimination.

[5] It is necessary to distinguish durability or quality (α) from lifetime or replacement period (T) if the two variables α and T are determined by different decision units. In investment theory, durability and lifetime are used interchangeably, because capital equipment is manufactured and used by the same decision unit, producers. In the case of consumer durables, α and T are distinct variables because producers and consumers may make different decisions regarding durability. Such a distinction becomes trivial only if α and T are related identically under varied market conditions.

[6] To be precise, P is also a function of Q, i.e., $P=P(\alpha, Q)$, where Q is the quantities demanded and supplied. This analysis has not explicitly incorporated Q into the model, because the choice of α and T is independent of equilibrium stock, Q, corresponding to a zero excess demand.

$$(1) \qquad P = P(\alpha) > 0 \qquad \text{for any } \alpha$$

and $P'(\alpha) > 0$.[7] For a diminishing-efficiency durable, maintenance and repair are required to prevent efficiency from falling as the durable ages. Let M represent the total operating expenses on maintenance and repair over the useful life T of the durable.[8] The present value of the total maintenance and repair expenses over T periods may be expressed as

$$(2) \qquad M = \int_0^T R(\alpha, t)e^{-rt}dt \qquad \text{for } T > 0$$

where $R(\alpha, t)$ denotes the operating expenses incurred by the durable of quality α in period t, and r is the rate of interest or discount; r is assumed to be constant throughout the useful life T of the durable.[9] Properties of maintenance and repair suggest that operating expenses increase with the age of the durable and are inversely related to the quality. Thus $R_t(\alpha, t) > 0$ and $R_\alpha(\alpha, t) < 0$. It is noted that $M = 0$ if and only if any of the following cases occur: (i) The good concerned is nondurable and T may be assumed to be zero. This case should be treated separately, because the service flow of durables comes into being only when $T > 0$. (ii) The consumer has not put the durable into service. This case may be excluded from the analysis because no replacement problem would arise. (iii) The durable concerned is of the one-hoss shay variety. Such a durable has a constant life span over which no repair and maintenance are required. In the one-hoss shay case, the durability problem can be solved by setting $\alpha = T$ and $M = 0$.

To simplify the analysis, it is assumed that the durable is maintained and repaired continuously so that it remains "as good as new" during its useful life. Provided that the intensity of usage is uniform over time, the durable will render service evenly over T periods. Some unit of measurement is selected so that the durable provides exactly one unit of service per unit of time, and the per unit expense of the service is defined as p. The present value of the total expense of the service over T periods is

$$(3) \qquad V = \int_0^T pe^{-rt}dt = \frac{p}{r}(1 - e^{-rT})$$

If scrap value is assumed to be nil, the total expense of the durable is composed of the price (P) and the operating cost (M). Thus,

$$(4) \qquad V = P + M$$

The consumer must compromise between replacement cost and operating cost by determining an optimum life of the durable. In a given planning horizon, the sooner the durable is replaced, the higher is the replacement cost, although the operating cost is lower. On the other hand, the operating cost would run higher should the durable be replaced too late. Taking α as a parameter, the consumer must determine the optimum value of T in the following expression:[10]

---

[7] All the functions used in this paper are assumed to be continuous and at least twice differentiable. Prime represents the operation of derivative, e.g., $P'(\alpha) = dP(\alpha)/d\alpha$ and $P''(\alpha) = d^2P(\alpha)/d\alpha^2$, and subscript represents the operation of partial derivative, e.g., $R_T(\alpha, T) = \partial R(\alpha, T)/\partial T$, $R_{TT}(\alpha, T) = \partial^2 R(\alpha, T)/\partial T^2$, and $R_{\alpha T}(\alpha, T) = \partial^2 R(\alpha, T)/\partial\alpha\partial T$.

[8] Parks' paper also discussed operating costs. Parks, however, confined his analysis to one-hoss shay durables, assuming that the amount of the variable input per unit of service flow remains constant over time and is increasing with durability. His operating costs, as illustrated by electricity bills incurred in the operation of lightbulbs, are different from the costs presented here. In this analysis operating costs other than maintenance and repair are assumed to be independent of durability and will be abstracted.

[9] Even though r varies continuously over time, r can be approximated by some constant. The mean value theorem of integral calculus makes such an approximation possible.

[10] The consumer's objective function chosen by

$$(5)\quad p = \frac{r}{(1 - e^{-rt})} \cdot \left[P(\alpha) + \int_0^{\mathrm{T}} R(\alpha, \mathrm{t}) e^{-r\mathrm{t}} d\mathrm{t} \right]$$

By minimizing p in equation (5), the optimal value of T, assumed to be in the interior,[11] can be obtained from the following conditions:

$$(6)\qquad p = R(\alpha, \mathrm{T})$$

$$(7)\qquad R_{\mathrm{T}}(\alpha, \mathrm{T}) > 0$$

The first-order condition for minimum p is given by equation (6), which shows that the consumer will not replace the durable until the unit expense of the service is equal to the marginal operating cost.[12] Equation (7) is the second-order condition, which is assured by an inherent property of repair and maintenance that $R_{\mathrm{t}}(\alpha, \mathrm{t}) > 0$. By substituting the optimum value of T into equation (5), the minimum value of p is obtained. The optimum values of p and T are denoted by p^* and T*. If $p^* > R(\alpha, \mathrm{T})$, the durable should be repaired and kept in service. Conversely if $p^* \leq R(\alpha, \mathrm{T})$ the durable should be discarded. Suppose for the durable concerned $p^* > R(\alpha, \mathrm{T})$ in the initial period. If the operating cost increases with T, then there must exist some T* such that $p^* = R(\alpha, \mathrm{T}^*)$.

Let p^c and p^m be the unit expense of the service incurred by consumers under perfect competition and pure monopoly, respectively. Then equation (6) may be written,

$$(6\mathrm{a})\qquad p^c = R(\alpha, \mathrm{T})$$

$$(6\mathrm{b})\qquad p^m = R(\alpha, \mathrm{T})$$

under two different market conditions. If $p^c < p^m$ as usually asserted, and if the competitive firms choose the same degree of durability α as the monopolist, then T in (6a) should be smaller than T in (6b), i.e., $\mathrm{T}^c < \mathrm{T}^m$. In other words, the durable sold by the competitive firms would be dis-

---

Brems (1968, pp. 37–41) may be expressed as $p = P/\mathrm{T} + nm/\mathrm{T}$ where m = average price of repairs, $n = \mathrm{T}^a/\alpha$ = total number of repairs over the useful life of the durable, and a = elasticity of total number of repairs $= d\,log(n)/d\,log(\mathrm{T})$. Assume that $e^{-r\mathrm{T}} \approx 1 - r\mathrm{T}$. Equation (5) in the present paper becomes

$$p \approx \frac{P(\alpha)}{\mathrm{T}} + \frac{\int_0^{\mathrm{T}} R(\alpha, \mathrm{t}) e^{-r\mathrm{t}} d\mathrm{t}}{\mathrm{T}}$$

which, except for the discount factor, is equivalent to Brems' objective function. Furthermore, setting $\alpha = \mathrm{T}$ in equation (5) yields $p = rP/(1 - e^{-r\mathrm{T}}) + R(\mathrm{T})$ which bears a resemblance to Parks' equation (24). See Parks, p. 44. If the good concerned does not deteriorate and there is no scrap value at the end of its lifespan, Schmalensee's equation (5) becomes $p = rP/(1 - e^{-r\mathrm{T}}) + m$. The rate of maintenance expenditure m is implicitly assumed to be independent of t or T. The consumer's decision variable, which does not play any role in Parks' equation (24), is identified as m and T in Schmalensee's and my equations, respectively. It should be noted that Schmalensee has abstracted T from his analysis, and that his model holds for any value of T including $\mathrm{T} = \infty$.

[11] The optimum value of T should fall somewhere between 0 and ∞, because (i) $P(\alpha) > 0$ for any economic good; (ii) $P(\alpha)/(1 - e^{-r\mathrm{T}})$ decreases as T increases, and (iii) $\int_0^{\mathrm{T}} R(\alpha, \mathrm{t}) e^{-r\mathrm{t}} dt/(1 - e^{-r\mathrm{T}})$ is increasing with T. Proofs of (ii) and (iii) are given as follows:

$$d[P(\alpha)/(1 - e^{-r\mathrm{T}})]/d\mathrm{T} = -rP(\alpha)e^{-rT}/(1 - e^{-r\mathrm{T}})^2 < 0$$

and

$$\begin{aligned} &d\left[\int_0^{\mathrm{T}} R(\alpha, \mathrm{t}) e^{-r\mathrm{t}}/(1 - e^{-r\mathrm{T}})\right] \Big/ d\mathrm{t} \\ &= \left[R(\alpha, \mathrm{T}) e^{-r\mathrm{T}}(1 - e^{-r\mathrm{T}}) - re^{-r\mathrm{T}} \int_0^{\mathrm{T}} R(\alpha, \mathrm{t}) e^{-r\mathrm{t}} d\mathrm{t}\right] \\ &\quad /(1 - e^{-r\mathrm{T}})^2 \\ &= \left\{re^{-r\mathrm{T}} \int_0^{\mathrm{T}} [R(\alpha, \mathrm{T}) - R(\alpha, \mathrm{t})] e^{-r\mathrm{t}} d\mathrm{t}\right\} \\ &\quad /(1 - e^{-r\mathrm{T}})^2 > 0 \end{aligned}$$

[12] It is interesting to note that the optimum useful life of the durable derived by Brems is merely a special case of the result provided in (6). This can be shown. The first-order condition for a minimum p for Brems' objective function (see fn. 10) is

$$\frac{-P}{\mathrm{T}^2} + \frac{m(a - 1)}{\alpha} \mathrm{T}^{a-2} = 0$$

which can be arranged into

$$\frac{-1}{\mathrm{T}} \left(\frac{P}{\mathrm{T}} + \frac{m}{\alpha} \mathrm{T}^{a-1} \right) + \frac{ma}{\alpha} \mathrm{T}^{a-2} = 0$$

Substituting p for the terms in parenthesis yields $p = d(nm)/d\mathrm{T}$, because $d(nm)/d\mathrm{T} = ma\mathrm{T}^{a-1}/\alpha$.

carded by the consumers sooner than the same durable sold by the monopolist.

Furthermore, it is observed that if the consumer minimizes p and replaces the durable every T^* periods, he will minimize the total present expense of the durable in an infinite economic horizon.[13]

Finally, a hypothetical question is raised concerning the consumer's choice of durability α. If the consumer is offered an opportunity to choose the quality of the durable, what choice will he make? Based on the principle of minimizing p in (5), the consumer would choose such α as to satisfy the following condition:[14]

$$P'(\alpha) = -\int_0^T R_\alpha(\alpha, t)e^{-rt}dt \tag{8}$$

Equation (8) implies that the increased (decreased) operating cost of a poorer (better) durable should be just offset by its lower (higher) price.

II. The Supply Side of the Durability Problems

In order to compare the supply and demand sides of the durability problem, the costs of the service rendered by the durable must be disaggregated into two parts: production cost and operating cost.[15] The production cost is the cost of producing a new unit of the durable, and the operating cost is the cost of maintaining and repairing an old unit of the durable. Assume the operating cost to be the same for the durable of the same age whether it be under the sales or rental policies. In the case that firms adopt the sales policy, the operating cost will be incurred directly by consumers rather than by producers as in the case of rentals. With the assumptions of identical production structure and constant returns to scale, an identical cost function is obtained for both monopolies and competitive firms. Let C be the production cost of a unit of the durable. Assuming C to be a monotonic increasing function of α, we obtain

$$C = C(\alpha) \tag{9}$$

Under the rental policy, the firm's objective function may be written

$$\pi = \frac{p^s}{r} - \frac{C(\alpha)}{1 - e^{-rT}} - \frac{\int_0^T R(\alpha, t)e^{-rt}dt}{1 - e^{-rT}} \tag{10}$$

where p^s is the seller's price per unit of the services as distinguished from p which may be referred to as the buyer's price. Let Q be units of the durable demanded or supplied. The buyer's price, which is the price acceptable to the consumers, is the function of Q and T as shown in equation (5),[16] while the seller's price, which is the price offered by the producers, is the function of Q only. It is assumed that for any values of α and T there exists some Q such that $p=p^s$. The value of Q so determined is the amount of the durable demanded and supplied in the stationary state of long-run equilibrium. Since the choice of α and T is independent of Q, this paper has excluded Q from the analysis. Such an omission obviously has no effect on the re-

---

[13] Equation (5) may be expressed in the following form:

$$p = r\left[P(\alpha) + \int_0^T R(\alpha, t)e^{-rt}dt\right] \cdot (1 + e^{-rT} + e^{-2rT} + \dots)$$

In the above equation p is minimized if and only if the total present cost of the durable in an infinite horizon attains the minimum.

[14] For one-hoss shay durables, the first-order condition for optimum quality is obtained as follows: $P'(T)/P(T) = r/e^{rT}-1)$. If perfect competition is assumed or if demand and cost functions are homogeneous of the same order with respect to T, the above condition will be identical to Swan's. See Swan (1970), p. 886.

[15] Smith has made a similar classification by disaggregating the cost of durable equipment into the cost on capital account and the cost on current account.

[16] Although Q does not appear in equation (5) for the reason explained in fn. 6, the proposition of demand theory that p is a function of Q remains valid.

sult of durability choice as long as stationary equilibrium is maintained. The variable π is the present value of the total net profit in the infinite planning horizon. It is assumed that all the durables have the same useful life T and that production of the durable recurs every T periods. By choosing the proper values of α and T, the firm's objective is to maximize π in (10). The first-order conditions for maximization are obtained as follows:[17]

$$(11)\qquad C'(\alpha) = -\int_0^T R_\alpha(\alpha, t)e^{-rt}dt$$

$$(12)\qquad R(\alpha, T) = \frac{r[C(\alpha) + M]}{1 - e^{-rt}}$$

Equation (11) shows that when the quality of the product is raised (lowered), the increased (decreased) production cost is just offset by the decreased (increased) operating cost. It is similar to the condition for the maximization of net benefit that requires marginal cost be equal to marginal benefit. The left-hand term in equation (12) is the marginal cost of rendering a unit of service, and the right-hand term is the corresponding average cost. The average cost is tantamount to interest payment per unit of time on the total cost of durable goods invested in the infinite horizon. A firm, when adopting the rental policy and investing one unit of the durable every T periods, will incur a total cost of $[C(\alpha)+M]/(1-e^{-rT})$. Accordingly, the right-hand term in equation (12) could be interpreted as the service charge per unit of time paid by the firm to the bank or bondholders. Thus, equation (12) states the equality of marginal and average costs as implied by minimization of the total cost.

It is observed from equations (11) and (12) that the optimum values of α and T are not related to the price or demand side. Both monopolies and competitive firms by maximizing net profit when adopting a rental policy would minimize the cost of the provision of services at the same level of durability. This result is essentially the same as Swan's proposition. Swan has implied, however, that his result applies to the case of sales policies.

Although under rental policies no consumer choice may exist regarding durability, it is instructive to compare the firm's choice of α and T with the values of α and T desired by the consumers. This can be done by comparing (11) and (12) with (8) and (6). Under competition a zero net profit in long-run equilibrium implies that $C(\alpha)=P(\alpha)$, and that $C'(\alpha)=P'(\alpha)$ for any α. As a result, (11) is identical to (8), and (12) is identical to (6). Therefore, under perfect competition, the choice of α and T by the firms and by the consumers are identical. In case of a monopoly, however, it may be assumed that $P(\alpha)>C(\alpha)$ for any α. Hence, $R(\alpha, T)<p$ via (5) and (12). The implication is that the monopolist would put the given durable in service for a shorter period than the rational consumer would prefer.

Next, attention is called to the sales policy under which the consumers determine the value of T. The firm must take the consumer's decision into consideration and maximize π in the following expression:

$$(13)\qquad \pi = \frac{P(\alpha) - C(\alpha)}{1 - e^{-rT}} + \lambda[p^s - R(\alpha, T)]$$

where λ is a Lagrange multiplier. Since a perfectly foresighted firm would always sell the durable at the price given in the following identity:

$$(14)\qquad P(\alpha) \equiv \frac{p^s}{r}(1-e^{-rT}) - \int_0^T R(\alpha, t)e^{-rt}dt$$

---

[17] The first-order condition for an optimum quality of the one-hoss shay is $C'(T)/C(T)=r/e^{rT}-1$. By comparing this condition with that provided in fn. 14, we conclude that the choice of durability by competitive firms accords with the consumer's.

$$(16)\qquad 0 = \frac{C'(\alpha) + \int_0^{\mathrm{T}} R_\alpha(\alpha, \mathrm{t})e^{-r\mathrm{t}}d\mathrm{t}}{1 - e^{-r\mathrm{T}}} + \lambda R_\alpha(\alpha, \mathrm{T})$$

$$(17)\qquad 0 = \frac{R(\alpha, \mathrm{T})(1 - e^{-r\mathrm{T}})e^{-r\mathrm{T}} - \left[C(\alpha) + \int_0^{\mathrm{T}} R(\alpha, \mathrm{t})e^{-r\mathrm{t}}d\mathrm{t}\right] re^{-r\mathrm{T}}}{(1 - e^{-r\mathrm{T}})^2} + \lambda R_{\mathrm{T}}(\alpha, \mathrm{T})$$

$$(18)\qquad p = R(\alpha, \mathrm{T})$$

equation (13) becomes

$$(15)\qquad \pi = \frac{p^s}{r} - \frac{C(\alpha) + \int_0^{\mathrm{T}} R(\alpha, \mathrm{t})e^{-r\mathrm{t}}d\mathrm{t}}{1 - e^{-r\mathrm{T}}} + \lambda[p^s - R(\alpha, \mathrm{T})]$$

Substituting $P(\alpha)$ in equation (5) by equation (14) yields $p = p^s$. Its explicit form is $p(Q, \mathrm{T}) = p^s(Q)$. Thus, Q can be solved as soon as the optimum value of T is determined. The first-order conditions for maximum π in (15) are given by equations (16)–(18). Through equation (18) the firm has taken into account explicitly the consumer's decision regarding the optimum value of T. If equations (16), (17), and (18) are consistent and independent, the three unknowns α, T, and λ can be solved.

Let us examine first the solution for the competitive case. By the nature of perfect competition, the rules of free entry and survival of the fittest would drive a firm's average revenue precisely to the level of average cost in long-run equilibrium. That is, $P(\alpha) = C(\alpha)$ for any α. Hence, equation (17) can be arranged as

$$(19)\qquad \frac{[R(\alpha, \mathrm{T}) - p]e^{-r\mathrm{T}}}{1 - e^{-r\mathrm{T}}} + \lambda R_{\mathrm{T}}(\alpha, \mathrm{T}) = 0$$

Since the first term on the left-hand side of equation (19) is zero by virtue of equation (18), and $R_{\mathrm{t}}(\alpha, \mathrm{t}) > 0$ for any $\mathrm{t} > 0$, it follows immediately that $\lambda = 0$. The implications are two-fold: 1) Under perfect competition each individual firm finds it impossible to improve its profit by changing the price of the durable or service. In a stationary equilibrium, a competitive firm can make only a zero profit, whether or not there is any consumer's constraint. 2) In a competitive market, producers and consumers make the same choice regarding durability, and thereby the same optimum values of α and T result from the sales and rental policies. Under the sales policy the optimum values of α and T are determined by equations (16) and (18)[18] while under the rental policy they are determined by equations (11) and (12). By making a comparison, equations (11) and (16) are both equivalent to equation (8), and equations (12) and (18) are both equivalent to equation (6). Recall that equations (6) and (8) provide the consumer's choice of α and T.

Under perfect competition, equations (17) and (18) are identical because $\lambda = 0$. This is not the case with a monopoly. Making use of equations (5) and (18), equation (17) can be expressed as

$$(20)\qquad \lambda = \frac{-[P(\alpha) - C(\alpha)]re^{-r\mathrm{T}}}{R_{\mathrm{T}}(\alpha, \mathrm{T})(1 - e^{-r\mathrm{T}})^2}$$

where λ cannot be zero because $P(\alpha) > C(\alpha)$ under monopoly. Substituting (20) in (16) yields:

---

[18] The Lagrange multiplier $\lambda = 0$ does not imply that the constraint is ineffective. See Zangwill, p. 42.

$$(21)\quad C'(\alpha) - \frac{rR_\alpha(\alpha, \mathrm{T})[P(\alpha) - C(\alpha)]}{R_\mathrm{T}(\alpha, \mathrm{T})(e^{r\mathrm{T}} - 1)} = -\int_0^\mathrm{T} R_\alpha(\alpha, \mathrm{t})e^{-r\mathrm{t}}dt$$

As a result of durability variation, consumers would adjust their choice of T, which in turn would bring about a change in net profit. Such a change in net profit is measured by the second term on the left-hand side of equation (21).[19] This is a major factor in the difference between the rental and sales policies. Under the rental policy, consumers cannot adjust the useful life of the durable according to the cost minimization principle which equates average expense with marginal expense. No difference is made under perfect competition, however, because the equality of average expense and marginal expense is assured by the fact that the price of the durable is always set equal to the minimum cost.

When equations (11) and (12) are compared with equations (21) and (18), it is obvious that the optimum solution for α and T under the rental policy may differ from that under the sales policy managed by the monopolist. Note that the right-hand side of equation (12) can be shown to be equal to p^c, and that under monopoly p in equation (18) should be taken as p^m.

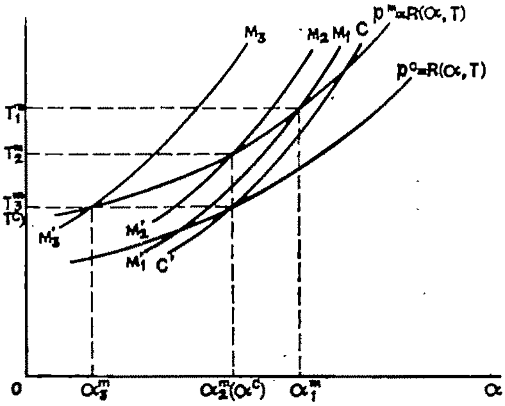


FIGURE 1

In Figure 1 equations (11) and (12) are plotted along with equations (18) and (21). The curve CC' on the $\alpha-\mathrm{T}$ plane represents equation (11), and any of the three curves M_1M_1', M_2M_2', and M_3M_3' may represent equation (21). It can be shown that the slope of CC', being positive, is greater than that of p^c or p^m,[20] and that all three curves M_1M_1', M_2M_2', and M_3M_3' are located to the left side of CC'.[21] The points α_1^m, α_2^m, and α_3^m represent the three possible values of durability chosen by the monopolist adopting the sales policy. As observed, $\alpha_1^m > \alpha^c$, $\alpha_2^m = \alpha^c$, and $\alpha_3^m < \alpha^c$. The point α^c is the optimum durability selected by the competitive firms under both rental and

---

[19] By virtue of $d\mathrm{T}/d\alpha = -R_\alpha(\alpha, \mathrm{T})/R_\mathrm{T}(\alpha, \mathrm{T})$ and $e^{r\mathrm{T}} \approx 1 + r\mathrm{T}$, the second term on the left-hand side of (21) may be expressed as $(d\mathrm{T}/d\alpha)(P-C)/\mathrm{T}$.

[20] Suppose that an interior maximum for π in (10) exists at some α and T. The second-order condition for the maximum is

$$\left[\frac{\partial^2\pi}{\partial\alpha^2}\right]\left[\frac{\partial^2\pi}{\partial\mathrm{T}^2}\right] > \left[\frac{\partial^2\pi}{\partial\alpha\partial\mathrm{T}}\right]^2$$

which gives

$$R_\mathrm{T}(\alpha, \mathrm{T})\left[C''(\alpha) + \int_0^\mathrm{T} R_{\alpha\alpha}(\alpha, \mathrm{t})e^{-r\mathrm{t}}d\mathrm{t}\right] > [R_\alpha(\alpha, \mathrm{T})]^2 e^{-r\mathrm{T}}$$

Taking the derivatives of (11) and (12) with respect to α and subtracting $d\mathrm{T}/d\alpha$ in (11) from that in (12) yields

$$\frac{[R_\alpha(\alpha,\mathrm{T})]^2 e^{-r\mathrm{T}} - R_\mathrm{T}(\alpha,\mathrm{T})\left[C''(\alpha) + \int_0^\mathrm{T} R_{\alpha\alpha}(\alpha, \mathrm{t})e^{-r\mathrm{t}}d\mathrm{t}\right]}{R_\alpha(\alpha, \mathrm{T})R_\mathrm{T}(\alpha, \mathrm{T})e^{-r\mathrm{T}}} > 0$$

Note that the slopes of the curves in Figure 1 are all positive because $d\mathrm{T}/d\alpha > 0$.

[21] Since $-rR_\alpha(\alpha, \mathrm{T})[P(\alpha) - C(\alpha)]/[R_\mathrm{T}(\alpha, \mathrm{T})(e^{r\mathrm{T}} - 1)]$ is positive whatever the values of α and T, it follows from equation (21) that $C'(\alpha) < -\int_0^\mathrm{T} R_\alpha(\alpha, \mathrm{t})e^{-r\mathrm{t}}d\mathrm{t}$. In contrast, equation (11) shows that $C'(\alpha) = -\int_0^\mathrm{T} R_\alpha(\alpha, \mathrm{t}) \cdot e^{-r\mathrm{t}}d\mathrm{t}$. Since the right-hand term of (11) increases with T as observed from

$$\partial\left[-\int_0^\mathrm{T} R_\alpha(\alpha, \mathrm{t})e^{-r\mathrm{t}}d\mathrm{t}\right] \Big/ \partial\mathrm{T} = -R_\alpha(\alpha, \mathrm{T})e^{-r\mathrm{T}} > 0$$

the curve of (21) should be located above the curve of (11).

sales policies. Furthermore, it follows from the earlier analysis that the monopolist adopting the rental policy would also choose durability α^c. However, it is found that the two situations $\alpha^m=\alpha^c$ and $T^m=T^c$ cannot coexist. When $T^m=T^c$, $\alpha^m<\alpha^c$; and when $\alpha^m=\alpha^c$, $T^m>T^c$. Swan's conclusion, represented by the limited case where $\alpha^m=\alpha^c$, implies that $T^m>T^c$.

It is suggested that the present analysis may be extended to study the choice of durability by a firm under a regulated monopoly environment.[22] What will be the optimum choice of durability if a monopolist is compelled to sell as well as rent his product? And what if a monopolist acts upon a court ruling to divest himself of a portion of his operations? It is implied in this study that a change in price or market structure brought about by regulation will have some effect on the choice of durability. However, the nature of change in durability level cannot be determined unless some explicit form of replacement and operating cost function is given.

III. Summary and Conclusion

In this paper it has been argued that the durability choice of the monopolist and the consumer cannot possibly be the same. Although such an inconsistency appears to be irrelevant to the durability analysis in those cases involving only the one-hoss shay goods or a rental-only policy, it may affect the optimum durability of the diminishing efficiency goods which are made solely for sales. The assumption of a perfect capital market would not resolve the inconsistency.[23] The major reason is that the consumer who bases his decision on the "price" or "expense" of the product would not put the durable in service for the same length of time as would the producers who base their decisions on the "cost" of the product. Moreover, it follows from the analysis that the monopolist's net profit should be lower under the sales policy than under the rental policy due to the effective constraint imposed by the consumers (i.e., $\lambda\neq 0$). Therefore, the monopolist would be inclined to rent the durable. Provided that the monopolist is forced to sell, the monopolist's choice of durability is *likely* to differ from that of competitive firms.[24] Under perfect competition, however, both consumers and producers would make the same choice and would be indifferent to the alternative policies.

---

[22] A referee has indicated that IBM agreed to a consent decree back in the 1950's to sell as well as rent its equipment and that the Bell System was forced by a FCC ruling to allow the attachment of privately owned PBX equipment to its lines. No doubt many interesting problems will arise from the analysis of such cases.

[23] The assumption of a perfect capital market is essential to the theory of optimum durability proposed by Swan. The same assumption was made by Douglas and Goldman in a deterministic model concerning monopolistic behavior. Although Swan has incorporated consumer-utility functions and durability choice in the mathematical appendix of his recent paper (1972, pp. 582–84), his argument is concerned merely with the durability of one-hoss shay "vehicles."

[24] In contrast with Schmalensee's result, this study does not preclude the case of $\alpha^m=\alpha^c$.

REFERENCES

H. Brems, *Product Equilibrium Under Monopolistic Competition*, Cambridge 1951, 30–36.

———, "Optimum Replacement of Consumer Durables," in *Quantitative Economic Theory: A Synthetic Approach*, New York 1968, 37–41.

A. J. Douglas and S. M. Goldman, "Monopolistic Behavior in a Market for Durable Goods," *J. Polit. Econ.*, Jan. 1969, *77*, 49–59.

E. Kleiman and T. Ophir, "The Durability of Durable Goods," *Rev. Econ. Stud.*, Apr. 1966, *33*, 165–78.

D. Levhari and T. N. Srinivasan, "Durability of Consumption Goods: Competition Versus Monopoly," *Amer. Econ. Rev.*, Mar. 1969, *59*, 102–07.

F. Lutz and V. Lutz, *The Theory of Investment of the Firm*, Princeton 1951.

D. D. Martin, "Monopoly Power and the Durability of Durable Goods," *Southern Econ. J.*, Jan. 1962, *28*, 271–77.

H. L. Miller, Jr., "On the Theory of Demand for Consumer Durables," *Southern Econ. J.*, Apr. 1961, *27*, 298–304.

R. W. Parks, "The Demand and Supply of Durable Goods and Durability," *Amer. Econ. Rev.*, Mar. 1974, *64*, 37–55.

R. L. Schmalensee, "Regulation and the Durability of Goods," *Bell J. Econ.*, Spring 1970, *1*, 54–64.

———, "Market Structure, Durability, and Maintenance Effort," *Rev. Econ. Stud.*, forthcoming 1974.

E. Sieper and P. L. Swan, "Monopoly and Competition in the Market for Durable Goods," *Rev. Econ. Stud.*, July 1973, *40*, 331–51.

V. L. Smith, "The Theory of Investment and Production," *Quart. J. Econ.*, Feb. 1959, *73*, 61–87.

———, *Investment and Production: A Study in the Theory of the Capital-Using Enterprise,* Cambridge 1961.

P. L. Swan, "Durability of Consumption Goods," *Amer. Econ. Rev.*, Dec. 1970, *60,* 884–94.

———, "The Durability of Goods and Regulation of Monopoly," *Bell J. Econ.*, Spring 1971, *2*, 347–57.

———, "Optimum Durability, Second-Hand Markets, and Planning Obsolescence," *J. Polit. Econ.*, May/June 1972, *80,* 575–86.

K. Wicksell, "A Mathematical Analysis of Dr. Akerman's Problem," in *Lectures on Political Economy,* vol. 1, London 1934, 274–99.

W. I. Zangwill, *Nonlinear Programming: A Unified Approach,* Englewood Cliffs 1969.

Earnings, Productivity, and Changes in Employment Discrimination During the 1960's

By JOAN GUSTAFSON HAWORTH, JAMES GWARTNEY, AND CHARLES HAWORTH*

Historians may refer to the 1960's as the decade in which Americans made a frontal attack on racial discrimination. A far-reaching Federal Civil Rights Act was passed. States passed additional legislation designed to promote racial equality and eliminate discrimination—particularly employment discrimination. Because of the political activism of the decade, a detailed analysis of the economic impact of these political changes is of great importance. This paper seeks to undertake such an analysis concentrating on the interactions between changing relative earnings, discriminatory employment practices, and relative productivity.

A number of writers have emphasized that the relative economic status of nonwhite males improved during the 1960's.[1] These findings are usually based on aggregated data, which indicate an improvement in the nonwhite/white earnings ratio (*NWER*). For instance, the *NW/W* median earnings ratio for males, age 25–64, increased 17.4 percent during the 1960's. While the size of this economic gain is of significance, its source and structure is also of great importance. The relative gains of nonwhites could have been achieved through two different mechanisms: the productivity of the nonwhite labor force may have increased relative to whites, and/or the intensity of employment discrimination may have declined.

Our analysis will seek to isolate the significance of changes in both the intensity of labor market discrimination and productivity according to race. Our findings contain some surprises. Approximately one-half of the gain in the *NWER* is attributable to the exiting of older nonwhite workers with low relative earnings combined with the entry of younger, better-prepared nonwhites who have high relative earnings. This can be illustrated by the fact that the relative gains of nonwhites within age cohort groups were modest—approximately 6 percent during the decade.

The largest gains were made by youthful, better-educated blacks. While earnings parity had not yet been attained, by 1970 the earning payoffs for young males continuing in school beyond the tenth grade were similar regardless of race.

Other researchers have argued that since nonwhites were systematically excluded from the "best" jobs, they had little incentive to acquire education and skills,[2] and were caught in "a cycle of poverty." With few skills and little education, they will have low earnings. However our findings would seem to suggest a basis

---

* Department of economics, Florida State University. This research is part of a larger research project on minority employment opportunities and economic growth that is being conducted in the Institute for Social Research at Florida State University. We gratefully acknowledge a supportive grant from the Economic Development Administration. Of course, we alone are responsible for the content of the paper.

[1] For example, see Richard Freeman, Finis Welch, and Daniel Moynihan.

[2] This view has been articulated by Paul Siegel, Bennett Harrison, Ivor Berg, and Barbara Bergmann and Gerolyn Lyle.

for cautious optimism in this area. Changes during the 1960's have increased the relative incentives for young nonwhites to acquire additional education and skills making the cycle of poverty hypothesis considerably less applicable today than in 1960.

This paper is organized into nine sections. In Sections I through IV a model of income determination is developed and tested. Sections V to VII use cohort data to analyze the changes in earnings over time for comparable age groups, including the important effect of educational payoffs. Section VIII analyzes income changes by educational groups to determine which group received the largest income increase during the decade. Section IX summarizes the findings of the paper.

I. A Model of Income Determination

In this section a model of income determination is developed which provides information on the importance of seven factors that affect income in addition to race. The factors are incorporated as follows:

Race (1 if nonwhite);
Age—a set of dummy variables for males, age 14–24, 25–34, 45–54, 55–64, and 65 and over, with the 35–44 age grouping acting as the base);
Years of schooling;
Rural resident (1 if rural);
Southern resident (1 if South);
Annual hours worked;
Marital status—a set of dummy variables for married, spouse present (*SP*), and persons once married (divorced, separated, or widowed), with singles acting as the base group;
Veteran status (1 if nonveteran).

We expected that individual earnings would peak for males between 35 and 54. The young (under 35) and the old (over 65) were expected to have significantly lower earnings than the middle-aged groups. Schooling was expected to positively influence earnings. Other studies have shown that earnings are lower in the South than the North and, similarly, rural status has a negative impact on money earnings. Hours worked is expected to positively influence earnings. The impact of marital status is not quite so obvious, but one would expect males with spouse present to have higher earnings than single males because of the specialization in provision of monetary income that is often associated with male family responsibilities.[3] We also anticipated that veterans might acquire certain skills and/or be less likely than nonveterans to have physical or mental handicaps. Thus we expect nonveteran status to reduce earnings.

With this model of income determination we are able to distinguish between earnings which are a result of certain productivity characteristics[4] and earnings which are not explained by these factors. In many studies this unexplained portion of earnings is attributed to discrimination since it can not be attributed to a particular income enhancing characteristic. In this paper we recognize that the differential is a function of other excluded factors as well as discrimination. When racial differences in the excluded factors are held constant at two points in time, then changes in the adjusted relative earnings, which our model estimates, are indicative of changes in employment practices.

II. Empirical Results—Aggregate Earnings Function Approach

Initially, we tested the model with data on the total population—including both

---

[3] See Gwartney and Robert Stroup for evidence on the impact of marital status on earnings.

[4] The term "productivity characteristics" is used in a broad sense, including locational factors and family responsibilities in addition to other factors that are more typically thought to influence productivity.

TABLE 1—THE AGGREGATED EARNINGS FUNCTION AND THE SEPARATED EARNINGS FUNCTIONS OF WHITE AND NONWHITE MALES FOR 1959 AND 1969
(t-ratios in parentheses)

Independent Variables	Dependent Variable-Individual Earnings Aggregated Function		Whites Only		Nonwhites Only	
	1959	1969	1959	1969	1959	1969
Nonwhite	−1000	−1427	—	—	—	—
	(19.3)	(16.3)				
Age (16–24)	−2004	−3732	−2070	−3916	− 992	−1688
	(34.7)	(38.4)	(33.0)	(37.2)	(10.1)	(8.1)
Age (25–35)	− 913	−1753	− 969	−1856	− 343	− 787
	(21.6)	(23.0)	(21.1)	(22.6)	(4.6)	(4.7)
Age (45–54)	190	278	211	380	− 53	87
	(4.1)	(3.6)	(4.2)	(3.4)	(0.7)	(0.5)
Age (55–64)	94	− 2	129	8	− 185	− 469
	(1.7)	(0.0)	(2.2)	(0.1)	(1.9)	(2.3)
Age (65+)	− 520	−1503	− 520	−1566	− 405	−1434
	(6.0)	(9.8)	(5.6)	(9.5)	(2.6)	(4.3)
Schooling	274	518	296	546	92	269
	(60.4)	(66.0)	(59.5)	(63.9)	(12.4)	(16.3)
Rural	− 789	−1155	− 834	−1198	− 499	− 616
	(22.6)	(19.5)	(22.3)	(19.0)	(7.1)	(3.8)
South	− 566	− 792	− 544	− 753	−1017	−1590
	(16.7)	(14.1)	(14.6)	(12.4)	(17.6)	(13.1)
Hours Worked	1.01	1.80	1.05	1.86	0.56	1.34
	(41.4)	(41.7)	(39.3)	(39.2)	(14.2)	(13.7)
Married (*SP*)	1065	1844	1152	1899	388	1449
	(22.0)	(22.1)	(21.7)	(20.8)	(4.9)	(8.6)
Once Married	266	711	227	733	10	649
	(3.0)	(5.7)	(2.7)	(5.2)	(0.1)	(2.9)
Nonveteran	− 45	− 158	− 49	− 169	− 108	− 83
	(1.3)	(2.8)	(1.4)	(2.8)	(1.8)	(0.7)
Y-Intercept	− 328	−2029	− 738	−2448	1796	429
R^2	.33	.32	.31	.31	.33	.29

Source: U.S. Census of Population, Public Use Sample 1960 and 1970.

white and nonwhite males—from the 1960 and 1970 *U.S. Census of Population*, Public Use Sample.[5] These results are presented in Table 1, columns 1 and 2.[6] In 1959 the unadjusted white minus nonwhite earnings were \$5,131–\$2,960, a differential of \$2,171. Adjusting the earnings differential for the seven "productivity factors" of the model reduced the *W/NW* estimated differential to \$1,000. The adjusted *NWER* was 80.5 percent compared to 57.7 percent for the unadjusted ratio (see Aggregated Earnings Approach, Table 2). In 1969, adjusting for the seven variables reduced the white-nonwhite differential from \$2,859 to the \$1,427 indicated in Table 1, column 2. The adjustment process increased the *NWER* from 66.0 percent to 83.0 percent in 1969. Thus, utilizing this method of adjustment, the seven factors of our model accounted for approximately one-half of the total *W-NW* earnings differential for both 1959 and 1969.

---

[5] This data base includes all males age 14 and over except persons who had either zero earnings or zero hours worked. The total number of males included in 1969 was 40,578 of whom 3886 were nonwhite. In 1959, the data base was 34,737 including 3385 nonwhites. All of the data presented in this paper will be for males only.

[6] An appendix containing the regression results for several alternative models and all of the supportive equations for the summary results presented in the paper is available from the authors upon request.

The independent variables of the model all have the expected signs, their magnitudes appear to be reasonable, and 11 of the 13 variables are significant at the 99 percent confidence level in both years. The "earnings functions" for the two years are parallel, indicating that the independent variables had a similar impact on earnings during the two time periods.

The *NWER*, adjusted for the seven factors, increased from 80.5 in 1959 to 83.0 percent in 1969, a gain of 3.1 percent. But in dollar terms, the *W/NW* adjusted differential actually widened during the 1960's. In money terms, the differential was $427 greater in 1969 than 1959. Measured in 1969 dollars, the gap increased by $169.

III. Differences in Earnings Functions According to Color

After adjusting for the seven factors of our model, the total population data suggest there was a small change of only 3.1 percent in the *NWER* during the 1960's. But these findings may be misleading because they implicitly assume that both whites and nonwhites have the same earnings function. The work of other researchers indicates this has not been the case. Failure to compensate for possible changes or differences in the earnings functions according to color during the 1960's could bias our estimates of changes in the adjusted *NWER* during the decade.

In order to isolate the importance of differences in the earnings functions according to color, we estimated the functions separately. Table 1, columns 3–6, presents the estimated earnings function of males according to color for both 1959 and 1969. There are, of course, several differences in these earnings functions. Since the mean earnings of nonwhite males is lower than white males, the coefficients are usually smaller. The estimated payoff for a year of education in 1959 was three times greater for whites than nonwhites—$296 versus $92. In 1969 it was two times greater—$546 versus $269. The age coefficients show a different lifetime earnings profile for nonwhites than for whites. A comparison of the age coefficients shows that the earnings of white males peak at a later age than nonwhites. This result probably reflects the overrepresentation of white males in white-collar occupations where earnings often continue to rise throughout one's working life. This age-earnings pattern was present for both 1959 and 1969.

The impact of marital and rural status was similar in 1969 and 1959. According to color, in absolute terms, the size of the coefficients for the marital and rural status variables were smaller for nonwhites, but relative to total earnings the coefficients were similar according to color. For both 1959 and 1969, southern residence reduced the earnings of nonwhites in dollar terms by approximately twice the reduction for white males. This regional difference is consistent with the findings of other researchers and supports the hypothesis that a higher degree of discrimination exists in the South.

IV. Changes in the Adjusted *NWER*—Separate Earnings Approach

Having estimated the earnings functions according to color, it is now possible to construct an estimate of earnings for one color group given the mean values of the other. This, in turn, permits disaggregating the earnings differences of the color groups into two components. First, one can derive a hypothetical *NWER* assuming that both white and nonwhite males had identical "earnings related" characteristics (i.e., the same mean values for years of schooling, age, veteran status, and other independent variables) but still possessed their respective "earnings functions." This might be called the

TABLE 2—ESTIMATED CHANGES IN THE *NWER* BETWEEN 1959 AND 1969[a]

	1959	1969	Change
Actual *NWER*	57.7	66.0	+14.4
Hypothetical *NWER*			
Identical Productivity: Different Earnings Function			
Both had white mean productivity values	70.8	79.0	+11.6
Both had nonwhite mean productivity values	70.9	76.9	+ 8.5
Identical Earnings Functions: Different Productivity Mean Values			
Both had white function	81.3	86.0	+ 5.8
Both had nonwhite function	81.5	83.5	+ 2.5
Aggregate Earnings Approach: Identical Earnings Function			
Different Mean Values	80.5	83.0	+ 3.1

Source: See Table 1.
[a] Shown in percent.

"identical productivity" estimate.[7] It removes the impact of differences, according to color, in the "productivity variables" of our model, but the impact of racial differences in the earnings functions will remain.

This hypothetical identical productivity *NWER*, presented in Table 2, will be less than unity solely because nonwhites on average receive a lower payoff for the productivity characteristics of our model. Assuming nonwhites (as well as whites) had the white mean value for the productivity variables, but still possessed the nonwhite earnings function, the hypothetical nonwhite earnings would have been $3634. Thus in 1959, if both groups had possessed the white mean values for the productivity variables of this model, the estimated *NWER* would have been $3634/$5131, or 70.8 percent. Assuming that both groups possessed the nonwhite productivity characteristics in 1959, the estimated hypothetical identical productivity *NWER* was 70.9 percent.

For 1969, our estimates suggest that nonwhites of identical productivity would have received between 76.9 percent (using nonwhite productivity variables for both color earnings functions) and 79.0 percent (using white mean values) of white male earnings. Thus when comparing white and nonwhite males who are similar with regard to seven productivity characteristics the *NWER* increased by between 8.5 and 11.6 percent during the 1960's.

A second hypothetical *NWER* can be developed using an "identical earnings function" assumption. This estimate uses the existing differences in the mean values of the earnings related variables of the color groups but assumes identical earnings functions according to color. For this hypothetical measure, there are no differences according to race in the *payoffs* from

---

[7] Two hypothetical estimates can be derived. The "identical productivity" estimates are:

$$\frac{Y_n(P_w)}{Y_w(P_w)} \quad \text{and} \quad \frac{Y_n(P_n)}{Y_w(P_n)}$$

where Y represents the actual earnings function according to color and P the actual mean values for each earnings related characteristic (the independent variables). Whites and nonwhites are denoted by w and n, respectively. The numerator of the first estimate is the hypothetical earnings of nonwhites assuming they had their actual earnings function but the white productivity characteristics. This is divided by the actual earnings of whites yielding the first identical productivity estimate. For the second estimate, the denominator is the hypothetical earnings of whites assuming their actual earnings function but the nonwhite mean values for each of the productivity characteristics. The actual nonwhite earnings is divided by this estimate to derive a second identical productivity estimate.

possession of productivity characteristics, but the presence of existing productivity differences remain.

The identical earnings function estimates also appear in Table 2.[8] For example, in 1959 if nonwhite males had received the white earnings function payoffs they would have earned $4173, while whites actually received $5131 for a hypothetical *NWER* of 81.3 percent. Using the black earnings functions, the hypothetical *NWER* was almost identical—81.5 percent. In 1969, the comparable estimates were 83.5 and 86.0 percent, indicating a gain of between 2.5 and 5.6 percent. This suggests that there was an improvement in the distribution of productivity characteristics of the nonwhite labor force relative to whites during the 1960's.

Table 2 shows that nonwhites would benefit more from having the white earnings function than the white productivity characteristics. This was true in both 1959 and 1969. The higher identical productivity *NWER* estimates in 1969 reflect the fact that the nonwhite earnings function was more similar to the white in 1969 than in 1959.

Our efforts in this section to isolate changes in relative productivity from discriminatory employment practices still suffer from a major shortcoming. If we have excluded other variables that influence earnings, our identical productivity estimates may be biased. If the excluded variables are associated with our other independent variables, they will bias our identical earnings function estimates. We would like to be able to "hold more things constant" and we will attempt to accomplish this objective in the following section.

V. Cohort Analysis

In attempting to analyze the effects of improving employment opportunities and changes in labor force characteristics, there are several advantages of using cohort data—data for the same age group at different points in time. First, comparison problems associated with the lifecycle pattern of earnings or changes in the lifecycle pattern are minimized. Cohort analysis allows one to measure the earnings of both whites and nonwhites at a similar point in their cycle of lifetime earnings.

Second, cohort comparison enables the investigator to hold "earnings related variables" constant over time. The data allow us to adjust in both 1960 and 1970 for factors such as rural, marital, and veteran status, southern residence, years of schooling, and hours worked. But we are unable to adjust directly for other factors such as health, achievement, motivation, and job preferences that might influence the current white-nonwhite earnings differential. Although there are racial differences in these and other earnings related variables, there is little reason to expect that such differences for adult age cohorts will change significantly over time.[9]

A third advantage of using cohort data is that it will allow us to say something about *changes* in the white-nonwhite differential in employment opportunities and economic preparedness of successive age groupings. Does each age group have the same opportunity as their predecessors ten years earlier? Or have changes taken

---

[8] The identical earnings function estimates are:

$$\frac{Y_w(P_n)}{Y_w(P_w)} \text{ and } \frac{Y_n(P_n)}{Y_n(P_w)}$$

where the former estimate assumes that both whites and nonwhites received the white earnings function payoff. The latter estimate assumes that both groups received the nonwhite earnings function payoff.

[9] A sample cohort of persons is drawn from basically the *same population* at two different points in time. Therefore, since the sample cohort is representative of the same population, we can expect a constancy of racial differences for many variables that are not directly included in our model.

place over time that allow successive generations different opportunities? The cohort comparison is particularly well suited to analyzing these types of questions.

VI. Cohort Changes

Utilizing the Public Use Sample data, files were created for the following age groupings: males 25–34, 35–44, and 45–54 in 1960 and males 35–44, 45–54, and 55–64 in 1970. By pairing these age groupings, it is possible to compare three age cohorts of adults during the 1960's. During the decade, the *NWER* within cohort groupings increased between 5.1 and 7.3 percent as illustrated in Table 3. These gains were considerably smaller than the 14.4 percent increase in the *NWER* estimated for all males.

The separate earnings approach of Section IV was also used to adjust the *NWER* for differences in productivity and earnings function within cohorts. We followed the same procedure as outlined above for the separate earnings approach, adjusting for the same seven factors. If both whites and nonwhites were assumed to have the white productivity mean values, the estimated adjusted *NWERs* in 1959 were 73.7, 70.8, and 65.1 percent for males of ages 25–34, 35–44, and 45–54, respectively. By 1969 the parallel estimates for the same cohorts groupings were 77.9, 74.0, and 70.6 percent, an average gain of 6.0 percent during the decade. Changes in the identical productivity estimates reflect only relative changes in the earnings functions, because productivity is held constant. Since these estimates are for cohorts, we are also holding several other factors relatively constant, even though they are not directly included in our model. The results imply that nonwhites of similar productivity as whites found their relative earnings increasing by approximately 6.0 percent during the 1960's. This is lower than our previous estimate, using the separate earnings approach, when males of all age groups were included. It probably reflects the greater similarity of the excluded factors for the cohort groupings.

TABLE 3—ESTIMATED CHANGES IN THE *NWER* WITHIN COHORT GROUPINGS BETWEEN 1959 AND 1969[a]

Cohort Group	1959	1969	Change
Actual *NWER*			
25–34 in 1960 (35–44 in 1970)	60.6	65.0	+7.3
35–44 in 1960 (45–54 in 1970)	56.8	60.3	+6.2
45–54 in 1960 (54–64 in 1970)	53.2	55.9	+5.1
Unweighted Mean	56.9	60.4	+6.2
Hypothetical *NWER*—Separate Earnings Approach			
Identical Productivity (white)			
25–35 in 1960	73.7	77.9	+5.7
35–44 in 1960	70.8	74.0	+4.5
45–54 in 1960	65.1	70.6	+8.4
Unweighted Mean	69.9	74.1	+6.0
Identical Earnings Function (white)			
25–34 in 1960	80.4	83.3	+3.6
35–44 in 1960	76.3	75.6	−1.0
45–54 in 1960	69.7	72.5	+4.0
Unweighted Mean	75.5	77.1	+2.1

Source: See Table 1.

[a] Shown in percent.

The identical earnings function estimates reflecting only changes in the amount of productivity factors possessed are also presented in Table 3 according to cohort. As would be expected, within adult cohorts, the estimates indicate that relative changes in mean productivity levels were small. The *NWERs* using the unweighted means for the identical earnings function estimates were 75.5 percent in 1959 and 77.1 percent in 1969, a gain of only 2.1 during the decade.

Table 3 yields considerable insight into the mechanism that brought about the relative earnings gains of nonwhites during the 1960's. The increase in the *NWER* within cohorts during the 1960's was much smaller—less than half of the total aggregated gains of nonwhite males. And the *NWER*, both unadjusted and adjusted for relative productivity differences, is negatively associated with age. The younger the age grouping, the higher the *NWER*. This is an important factor in explaining the magnitude of the gains in the *NWER* for all males in the total population during the 1960's. Older nonwhites with the lowest earnings relative to whites of corresponding age exited from the labor force and were replaced by younger nonwhites, who are better prepared for the labor market. This mechanism, the exiting of the older group and the entry of the younger group, accounted for approximately one half of the 14.4 percent gain during the decade.

VII. Educational Payoffs Within Cohort Groups

Past researchers have noted racial differences in the earnings payoff derived from schooling.[10] The education-earnings relation is particularly important since it not only directly influences economic status, but it also will influence the incentive to acquire human capital. Table 4 summarizes the earning payoffs from schooling for cohort groups. Two aspects of the estimates stand out. First, the relative education earnings payoff is definitely negatively related to age.[11] The older age groupings have lower *NW/W* education-earnings payoffs. In both 1959 and 1969, the relative payoffs of nonwhites 45–54 in 1960 were less than half the payoff for their "sons" in the 25–34 (in 1960) grouping. Second, while there was wide variance in the education-earnings payoff among age groupings, there was little change *within a cohort* during the 1960's. For instance, the relative marginal earnings per unit of education for nonwhites 25–34 in 1960 was 53.4 percent of whites in 1959 and 59.5 percent in 1969.

TABLE 4—THE ESTIMATED INDEPENDENT IMPACT OF A YEAR OF SCHOOLING ON THE EARNINGS OF WHITE AND NONWHITE MALES ACCORDING TO COHORT AND YEAR

	Estimated Change in Earnings Per Year of Schooling[a]		
Year and Cohort	Nonwhite[b]	White[b]	*NW/W*[c]
1959			
25–34 in 1960	$137	$257	53.3
35–44 in 1960	182	463	39.3
45–54 in 1960	104	444	21.1
1969			
25–34 in 1960	$388	$649	59.8
35–44 in 1960	320	741	43.2
45–54 in 1960	178	636	28.0

Source: See Table 1.

[a] These estimates of the coefficient of the schooling variable are for a model that includes rural status, southern residence, veteran status, marital status, and hours worked in addition to the schooling variable. The model was run separately according to color and age group for both 1959 and 1969.

[b] Shown in 1969 dollars.

[c] Shown in percent.

This result indicates that each successive cohort of nonwhites receives a higher earnings payoff from schooling than their predecessors. However, once the cohort is in the labor market there is little change in the relative payoff. There are a number of factors that might help explain this phenomenon. First, if achievement differences contribute to the lower educational payoff of nonwhites, the pattern of the payoffs over time implies that these differences according to color are declining. Perhaps this reflects the relative improvement in the quality of black schooling. Second, older nonwhites may have a lower earnings-education payoff (and a lower *NWER*) because of greater discrimination in past employment practices. The nonwhite high school graduate who was channeled into menial, low skill jobs by the employment practices of 10 to 20 years ago can never really make up this loss, even if current practices are less discriminatory.

---

[10] For example, see Freeman, Welch, Bergmann and Lyle, Harrison, Giora Hanoch, Gwartney, Dave O'Neill, Leonard Weiss and Jeffrey Williamson, C. Haworth and David Rasmussen, and R. Weiss.

[11] This result is consistent with the findings of Finis Welch, who concluded that the earnings payoff from school was negatively related to experience.

Taken together the cohort analysis of relative earnings and the earnings-education payoff strongly imply that (a) there has been a substantial improvement in the relative labor market preparedness of successive cohorts of nonwhite males, and/or (b) the discrimination of past employment practices is continuing to exert a substantial impact on current earnings according to color.

VIII. Analysis According to Educational Groups

Analysis of the relative earnings of separated educational groupings may shed light on structural changes in employment opportunities. Others have argued that employment discrimination is strongest at the very top of the job hierarchy and thus employment discrimination is strongest against the best educated blacks.[12] This hypothesis implies that the adjusted *NWER* and relative marginal payoffs to education would be lowest for blacks with the most education.

Table 5 presents evidence on the adjusted *NWER* within the three educational groupings by region and for younger persons. In 1959, the adjusted *NWER* was highest for the least educated and lowest for those with the most education. This pattern of the adjusted *NWER* is highly consistent with the "pattern of employment discrimination" theory.[13] In 1969, this was much less the case. During the 1960's, the largest gains in the adjusted *NWER* were recorded by the best educated. In the North, the adjusted *NWER* for males with 0–10 years of education increased by 1.8 percent. Similar small relative gains were recorded by the poorly educated in the South and by younger persons with little schooling. By contrast, the estimated adjusted *NWER* ratio for nonwhite males with 14 or more years of schooling increased by 15.3 percent in the North, and 13.9 percent in the South. The relative increase in the *NWER* for those with 11–13 years of schooling also exceeded those of the poorly educated group.

Table 5—The Adjusted *NWER* for Males with 0–10, 11–13, and 14+ Years of Schooling According to Region and for Persons Age 14–34 for 1959 and 1969

Group by Years of Schooling		Adjusted *NW/W* Earnings Ratio[a] 1959	1969	Change
Non-South	0–10	82.8	84.3	+ 1.8
	11–13	76.7	82.3	+ 7.3
	14+	72.4	82.5	+15.3
South	0–10	72.6	74.3	+ 2.3
	11–13	64.5	74.0	+14.7
	14+	61.0	69.5	+13.9
Persons: Ages 14–34				
	0–10	83.3	86.3	+ 3.6
	11–13	76.3	86.0	+12.7
	14+	77.3	85.3	+10.3

Source: See Table 1.

[a] The aggregated earnings approach was utilized to adjust for years of schooling, age, marital status, rural status, region, hours worked, and veteran status. Ratios shown in percent.

When only younger earners—persons age 14 to 34—are considered, the negative relation between education and the adjusted *NWER* had evaporated completely by 1969. The adjusted *NWER* for all three educational groupings was approximately 86 percent. These findings indicate that structural changes in employment opportunities during the decade reduced the strength of the pattern of discrimination hypothesis.

The *NW/W* marginal payoffs per year of schooling for persons age 14–34 within educational groupings are shown in the following tabulation for the estimated adjusted additional earnings per year of schooling:

---

[12] For example, see Harrison and Siegel.

[13] Of course, other factors such as a widening achievement gap as quantity of schooling increases may also explain this pattern of the *NWER*.

		Nonwhite	White	*NW/W*
1959	0–10	$ 49	$117	.42
	11–13	$122	$237	.51
	14+	$349	$133	2.62
1969	0–10	$ 43	$ 73	.59
	11–13	$376	$294	1.28
	14+	$362	$393	.92

Except for those with 0–10 years of schooling, the adjusted marginal payoffs from additional schooling were similar according to color in 1969. These data are also consistent with a weakening of the pattern of discrimination hypothesis during the 1960's. The pattern of discrimination theory has sometimes been used to argue that it was "rational" for nonwhites to become "school dropouts." The data presented in this section suggest that as of 1969, white and black males with 11 or more years of schooling had a similar monetary incentive to *continue* on to a high school or college degree.

IX. Summary and Conclusions

The aggregate increase in the *NWER* during the 1960's stemmed from two major sources. First, the *NWER* increased an estimated 6.0 percent because nonwhite males were able to improve the relative payoff derived from their productivity factors. This gain would reflect primarily improved employment opportunities and a decline in employment discrimination against blacks. In addition, approximately 2 percent is estimated to stem from gains in the amounts of productivity factors possessed by nonwhites relative to whites. A second major source of change—accounting for gains of between 6 and 8 percent in the *NWER* during the decade—was attributable to the exiting of older nonwhite workers with low relative earnings combined with the entry of younger, better-prepared nonwhites who have high relative earnings.

A consistent negative relationship exists between age and the *NW/W* education-earnings payoff. While successive cohorts of younger black males receive a greater relative payoff from their schooling, the estimated *NW/W* earnings per year of schooling within cohort groupings increased only slightly during the decade.

There is evidence that important structural changes in employment opportunities took place during the 1960's. The largest gains in relative earnings were recorded by those with the most education. The *NWER* for persons with 14 or more years of schooling increased 15.3 percent in the North and 13.9 percent in the South. A similar pattern of earnings gains was recorded by younger people. In addition, the relative payoffs from continuing in school were greater for the better educated. By 1970, the *additional* earnings per year of schooling for younger (under 35) nonwhite males with 11 or more years of schooling were very similar to whites. The disincentive of nonwhites to continue past the tenth grade caused by low marginal payoffs to education was sharply reduced during the decade. Thus, while the aggregated data overstate the earnings gains attributable to improved employment opportunities, they also fail to reveal important structural changes in relative earnings. The structure of the changes, for both the *NWER* and the payoffs from additional schooling, suggests that progress was made during the 1960's toward breaking the cycle of poverty resulting from discriminatory employment practices.

REFERENCES

I. Berg, *Education and Jobs: The Great Training Robbery*, New York 1970.

B. R. Bergmann and G. Lyle, "The Occupational Standing of Negroes by Areas and

Industries," *J. Hum. Resources*, Fall 1971, *6*, 411–33.

R. B. Freeman, "Decline of Labor Market Discrimination and Economic Analysis," *Amer. Econ. Rev. Proc.*, May 1973, *63*, 280–86.

J. Gwartney, "Discrimination, Achievement and Payoffs of a College Degree," *J. Hum. Resources*, Winter 1972, *7*, 60–70.

——— and R. Stroup, "Measurement of Employment Discrimination According to Sex," *Southern Econ. J.*, Apr. 1973, *39*, 575–87.

G. Hanoch, "An Economic Analysis of Earnings and Schooling," *J. Hum. Resources*, Summer 1967, *2*, 310–29.

B. Harrison, "Education and Underemployment in the Ghetto," *Amer. Econ. Rev.*, Dec. 1972, *62*, 796–812.

C. Haworth and D. Rasmussen, "Human Capital and Inter-Industry Wages in Manufacturing," *Rev. Econ. Statist.*, Nov. 1971, *53*, 376–80.

D. P. Moynihan, "The Schism in Black America," *Publ. Interest*, Spring 1972, *27*, 3–24.

D. O. O'Neill, "The Effects of Discrimination on Earnings: Evidence from Military Test Score Results," *J. Hum. Resources*, Fall 1970, *5*, 475–86.

P. M. Siegel, "On the Cost of Being a Negro," *Sociological Inquiry*, Winter 1965, *35*, 41–58.

L. Weiss and J. Williamson, "Black Education, Earnings and Interregional Migration: Some New Evidence," *Amer. Econ. Rev.*, June 1972, *62*, 372–83.

R. Weiss, "The Effect of Education on the Earnings of Blacks and Whites," *Rev. Econ. Statist.*, May 1970, *52*, 150–59.

F. Welch, "Black-White Differences in Returns to Schooling," *Amer. Econ. Rev.*, Dec. 1973, *63*, 893–907.

U.S. Census of Population, Public Use Sample, 1960 and 1970.

Direct Measurement of Popular Price Expectations

By GEORGE DE MENIL AND SURJIT S. BHALLA*

Expectations are an important determinant of economic behavior. The analysis of economic expectations, however, has suffered from inadequate measurement. Throughout the literature, expectations have been treated as unobservable, and have therefore been measured only indirectly. The object of this paper is to develop and test a direct measure of the expectations of inflation, a particularly critical economic variable.

A direct measure of inflationary expectations is here taken to be one constructed from a sample survey in which individuals are asked to state their expectations explicitly. The paper focuses on a popular forecast of the rate of change of consumer prices in the forthcoming year as calculated from responses to the *Survey of Consumer Finances* (*SCF*). This measure is presented in Section I.

The *SCF* and others like it have existed for some time, but, a few recent studies notwithstanding, most economists continue to regard survey measures as inherently unreliable.[1] Their principal criticism is that the opinions an individual expresses in an interview are not the same as the opinions on which he acts. At issue in this criticism is whether direct or indirect measures of expectations can best explain actual behavior. Tests of the explanatory power of the direct measure of expected inflation developed in Section I are presented in Sections II and III. The results demonstrate that the *SCF* measure of expected inflation, particularly when combined with a related measure of unexpected inflation, explains wage increases better than several currently used indirect measures. In particular, when the *SCF* measures are added to the wage equations of George Perry, Otto Eckstein and Roger Brinner, and Robert Gordon (1971), they perform better than these authors' own indirect measures in the first two cases, and as well as the indirect measure used in the third case.

I. Price Expectations in the *Survey of Consumer Finances*

The *Survey of Consumer Finances* is a survey of a representative sample of *U.S.* households conducted every year by the Survey Research Center of the University of Michigan. (See the Appendix for details.)

One of the questions asked in almost every survey has been the following:[2]

---

* Assistant professor of economics and graduate student, respectively, Princeton University. We are grateful for the comments of Robert Solow, Paul Wachtel, and the referee. Part of our research was supported by the National Science Foundation.

[1] The responses to questions regarding inflationary expectations in the *SCF* have not been widely used. The exceptions are the reports of the Michigan Survey Research Center and the consumption studies of Eva Mueller, Thomas Juster and Paul Wachtel (whose study in the *Brookings Papers* uses a measure similar to the one developed here), and Lester Taylor. One survey measure of price expectations which has attracted attention recently is the biannual survey of business economists conducted since 1947 by Joseph A. Livingston, one-time financial editor of the *Philadelphia Bulletin*. The Livingston measure has been analyzed by Stephen Turnovsky, and has been used in wage studies by Robert Gordon and by Turnovsky and Michael Wachter, and in interest rate studies by D. H. Pyle and William Gibson.

[2] The precise wording of the question has in fact varied from time to time, as is pointed out in the Appendix. The wording given here is that used after 1960. See *1970 Survey of Consumer Finances*, p. 257.

"Thinking about prices of things you buy in general, do you think they will go up in the next year or so, or go down, or stay where they are now?" In addition, since the second quarter of 1966, respondents expecting prices to go up have been asked: "How large a price increase do you expect? Of course nobody can know for sure, but would you say that a year from now prices will be about 1 or 2% higher, or 5%, or closer to 10% higher than now, or what?" It is possible to estimate the mean and other moments of the distribution of expectations of price change since 1966:2 from the answers to both of these questions. The critical problem, however, is how to infer similar quantitative estimates from the limited data available before 1966.

A suggestive discussion of this problem can be found in Henri Theil's analysis of the Munich Business Test in chapter 4 of *Economic Forecasts and Policy*. Theil constructs from the percentage of respondents expecting sales to rise, fall, and stay the same, a dispersion index, which he also calls the balance.

$$BAL = \frac{\%\ \text{Rise} - \%\ \text{Fall}}{\%\ \text{Rise} + \%\ \text{Fall} + \%\ \text{Same}}$$

He then discusses qualitatively the relationship between BAL and the complete distribution of forecasts. His discussion can be pushed further and applied to the problem of price expectations.

Let X be an individual respondent's expectation of the rate of price increase in the forthcoming year; let the density and cumulative distribution over the population of such expectations be, respectively, $f(X)$ and $F(X)$. Let the mean of this distribution be $EXPT$ and the standard deviation be SIG. If one takes the "stay the same" response to correspond to a prediction of some small change between $+\delta$ and $-\delta$, and if, for the sake of simplicity, one assumes these responses to be equally distributed between positive and negative changes, then

$$BAL = \int_0^\infty f(X)dX - \int_{-\infty}^0 f(X)dX$$

or

$$(1) \quad BAL = 1 - 2F(0)$$

If the shape of the distribution were known, (1) would provide a relation between BAL and $EXPT$ which could be used to estimate $EXPT$ prior to 1966:2.

The assumption maintained in this paper is that $f(X)$ is normal.[3] It then follows that

$$Z = \frac{X - EXPT}{SIG}$$

is a standard normal variate. If $G(Z)$ is its cumulative distribution then (1) may be written as

$$(1') \qquad BAL = 1 - 2G\left(\frac{-EXPT}{SIG}\right)$$

This is a relation between $EXPT/SIG$ and BAL, which, when plotted by means of the standard normal tables, is seen to pass through the origin and to be approximately linear over most of the range of BAL.[4]

The following linear approximation to (1′) was estimated using post-1966:2 data for the three relevant variables.

$$(2) \qquad \frac{EXPT}{SIG} = \underset{(140.3)}{1.50}\ BAL$$

---

[3] Unfortunately, the *SCF* data themselves provide no way to test this assumption. Answers to the quantitative question asked since 1966:2 were coded in five intervals (see Appendix). These answers, however, cluster in the three intervals 1–2 percent, 5 percent, and 10 percent or more, and this makes it impossible to use them to test for the shape of the underlying distribution.

[4] BAL is both positive and negative in the post-World War II period, but its absolute value is rarely greater than .85.

period of fit: 1966:2–1972:3,[5] with omissions

$$R^2 = .922 \qquad \text{D.W.} = 1.812$$

Equation (2) could permit backwards extrapolation of *EXPT* if *SIG* were known. But, like *EXPT* itself, *SIG* can only be measured directly after 1966:2. Therefore an attempt was made to derive an equation for *SIG* using the post-1966:2 data. The plot of *SIG* revealed a modest rise between 1966 and 1968 from 2.5 to 2.6 followed by a sharp increase to 3.0 in 1969 and 1970. A plausible interpretation of this pattern is that it reflects a tendency for the standard deviation to rise as the mean rises plus an increase in popular uncertainty about the future course of prices in the first two years of the Nixon Administration, when inflation and unemployment were both increasing at the same time. In order to test this hypothesis it was decided to regress *SIG* after 1966:2 against *EXPT* and an index of popular uncertainty about economic conditions. The index chosen, with some reservations, was the Index of Consumer Sentiment, an average of dispersion indices based on five attitudinal questions in the *SCF*. Though these questions ask about optimism and pessimism concerning economic conditions, there is also some basis for treating the Index as a measure of uncertainty.[6]

In line with the preceding discussion, the following equation was estimated.[7]

$$(3)\quad \underset{(17.1)}{SIG = 3.56} + \underset{(4.30)}{.20EXPT} - \underset{(-15.6)}{.018MOOD}$$

period of fit: 1966:3–1971:2, with omissions

$$R^2 = .968 \qquad \text{D.W.} = 1.89$$

MOOD is the Index of Consumer Sentiment. Combining (2) and (3) one obtains the following equation for backwards prediction of *EXPT*:[8]

$$(4)\quad EXPT = \frac{\underset{(17.0)}{5.35BAL} - \underset{(-15.6)}{.027MOOD\cdot BAL}}{1 - \underset{(4.27)}{.30BAL}}$$

This equation was used to calculate values for *EXPT* from 1952:4 through 1966:1, which were then combined with the measure of *EXPT* taken directly from the *SCF* from 1966:2 through 1972:3. Details of data construction are discussed in the Appendix.[9]

The series for *EXPT* thus constructed remains a crude measure of inflationary expectations. Because of its heavy reliance on the balance, which is only a summary statistic, the measure ignores some of the information contained in the responses, and in fact may be biased.[10] Despite these recognized drawbacks, the balance statistic has featured prominently in most empirical work on surveys to date. Moreover, even if it is biased, *EXPT* may still be a better measure of inflationary expectations than the indirect measures used in most empirical work. The next section reports the results of tests comparing *EXPT* with alternative indirect measures.

---

[5] *EXPT* and *SIG* were estimated from the quantitative estimates of expected inflation. The omissions refer to the quarters (1966:4, 1967:2, and 1968:1) for which such data were not available. (*T*-statistic in parentheses.)

[6] A recent important study on consumption by Juster and Wachtel (1972a, b) treats the Index as a measure of uncertainty. Also, Dwight Jaffee reported finding that the spread between BAA and AAA bonds of the same type of issuer and maturity was highly negatively correlated with the Index of Consumer Sentiment.

[7] In order to adjust for negative serial correlation, equation (3) was estimated using two-quarter averages of all variables. Regarding the omitted observations, see fn. 5. The sample period here stops short of the Phase 1 and Phase 2 controls of the Nixon Administration because of the presumption that the controls may have caused structural change in the degree of uncertainty associated with price expectations.

[8] Each coefficient in equation (4) is the product of coefficients in (2) and (3). Each variance was estimated as the variance of the linear terms of the Taylor's series approximation to the product. The *t*-statistic in parentheses under the coefficient is based on that estimated variance.

[9] It was felt that, the arguments in the text notwithstanding, some uncertainty remains regarding the proper interpretation of *MOOD*, and consequently regarding the validity of equation (3) and the related equation (4). Therefore, an alternative measure of *EXPT* was constructed based on the assumption that prior to 1966:2, *SIG* was a constant equal to its average value from 1966:2 to 1968:4. All of the experiments reported in Sections II and III were also conducted with this alternative measure. As it turned out, both of the series and the results of the tests based on them were very similar. When compared with its sharp movement in 1969–70, the entire previous history of *MOOD* back to the first observation in 1952:4 is relatively stable. Therefore the behavior of an *EXPT* series based on *MOOD* is quite similar to that of an *EXPT* series based on the assumption of a constant *SIG*.

II. Competing Measures

As explained earlier, it was decided to test the validity of *EXPT* by comparing its ability to explain wage increases with that of the indirect measures used in three recent equations which were the subject of a comparative study by Gordon (1972).[11] The equations—by Perry, Eckstein and Brinner, and Gordon (1971)—were selected because of the convenience of Gordon's comparative study and because in two of these three equations, particular emphasis had been placed on the specification of the price expectations variable.

In the three test equations, as in most work on economic expectations, expected price change is assumed to be a function of past rates of actual price change. Perry's specification of expected price change is the simplest, namely that it is equal to the quarterly rate of change of the consumer price index, lagged one quarter. Eckstein and Brinner lengthen somewhat the distributed lag on past price changes and add an additional threshold effect.[12] Their distributed lag spans the full year prior to the current quarter, and has weights which decline linearly to zero. Their threshold level is 2.5 percent annual inflation, and their additional expectation variable is the positive excess over the threshold of the annual rate of inflation during the two previous years. Gordon's specification is the most complex. His expectations variable is a twelve-quarters long, freely estimated, polynomial, distributed lag on the quarterly rate of change of the deflator for personal consumption expenditures. In addition, he introduces another price variable to represent the short-run effect of product demand on labor, namely an eight-quarter polynomial distributed lag function of the difference between the rate of change of the deflator for private nonfarm *GNP* and the personal consumption deflator. Conceptually this is not an expectations variable, but it is clear that statistically it may be correlated with expectations variables.

All three of these indirect representations of inflationary expectations are quite different from *EXPT*, as evidenced by their low correlations with *EXPT* (Perry, 0.69; Eckstein and Brinner, 0.59; Gordon, 0.60).[13]

The balance of the specification of each of the test equations is described in the studies cited above, including Gordon's comparative study.

---

[10] The balance statistic is constructed in such a way as to be roughly invariant to changes in the percentage of respondents saying "stay the same." A more refined method for inferring the mean expectation from responses to a question about the direction of change has been used by Harry Shufford, John Carlson and Michael Parkin, and de Menil. The statistical properties of the measure used in these studies have been analyzed in depth by Carlson and Harl Ryder.

[11] Gordon kindly supplied us with his complete data base.

[12] They also, as does Gordon, replace the consumer price index with the deflator for personal consumption expenditures.

[13] These correlations refer to the period 1954:1–1971:2. The Eckstein-Brinner measure is taken to be the weighted average of their distributed lag variable and their threshold variable with the weights based on the coefficients of these variables in the two-quarter estimate of their original equation.

Gordon's estimation procedure, which is followed here, is to estimate each equation in both one-quarter and two-quarter form. In the one-quarter form, the dependent variable is a one-quarter rate of change of the wage rate at an annual rate, and all independent variables are either one-quarter rates of change at an annual rate or one-quarter averages of level variables, or distributed lag functions of either of these. The two-quarter form is simply a two-quarter average of both the left- and right-hand sides of the one-quarter equation.

Gordon's estimates of the original three test equations for the period 1954:1–1970:4 appear in column 1 of Tables 1, 2, and 3.[14] In each block of the column, the two-quarter estimate and its t-statistic are presented above the corresponding one-quarter estimate. Precise variable definitions appear in the glossary accompanying each table.[15]

III. Significance Tests

Expected inflation influences wage change in a very simple manner in each of the three test equations. An increase in expected inflation results in a fractional increase in wage change in the following quarter. The availability of a direct measure such as ours permits a more complex modeling of the impact of expectations. One hypothesis, originally advanced by Stephen Turnovsky and Michael Wachter, is that in addition to the initial impact, there is a "catch-up" effect whereby wages catch up with a portion of the difference between recent inflation and previous expectations.

This effect is difficult to represent if one's measure of expectations is itself based on recent inflation. But with a direct measure like $EXPT$, the effect can readily be represented by the difference between recent inflation and an appropriate past value of $EXPT$. Reasoning that unexpected inflation influences wage change through increasing worker demands in new contract negotiations, and activating escalator clauses, we maintain that it is the average amount of unexpected inflation over the past year which is relevant[16] and, therefore, that the catch-up effect may be represented by $CATCH$, where

$$CATCH = \overset{4}{g_d} - \overline{EXPT}_{-4} \tag{5}$$

In this expression, g_d^4 is the four-quarter rate of change of the deflator for consumer expenditures d; $\overline{EXPT}_{-4}$ is the average of $EXPT$ four and three quarters ago. This average is used in place of $EXPT$ itself in order to remove some of the purely random variability of the series.

In view of the above, two tests were performed with each of the selected equations. The first was a simple comparison of

---

[14] Our estimates of the equation in col. 1 of Table 3 differ slightly from Gordon's because, unlike Gordon, we were unable to constrain the first derivatives of the polynomial lags to be zero in the furthest period. The differences, however, are small, and all our other benchmark results agree exactly with Gordon's.

[15] In addition to the variables listed in Tables 1, 2, and 3, each of the original equations has one or more different personal and/or social security tax variables. All of these variables are highly orthogonal to the other independent variables and their estimated coefficients are quite insensitive to changes in the other independent variables; therefore we have excluded these coefficients from our tables.

[16] The average duration of union contracts is about three years and has not changed much between 1954 and 1970. See Cornelia Ward (1959, 1965), and M. E. Sparrough and L. W. Bolton. There is less evidence on the frequency of wage adjustments in nonunion employment, but the norm appears to be one year. If union contracts are given a weight of .2 and nonunion wage adjustments a weight of .8, a rough estimate of 1.4 years is obtained for the average time span between wage decisions in the private economy. In order to allow for the existence of escalator clauses which can be activated quarterly, this estimate was adjusted downward to one year. The differences in the results with an alternative, six-quarter specification of the catch-up effect turned out to be marginal.

TABLE 1—PERRY WAGE EQUATION

Variable	Original Equation	*EXPT* Added	*EXPT* and *CATCH* Added	Perry's Expectations Excluded
Inverse of Weighted Unemployment, $1/U^*$	.0001	.0001	.0001	.0001
	(.42)	(.30)	(.27)	(.28)
	.0001	.0001	.0001	.0001
	(.46)	(.46)	(.44)	(.43)
Unemployment Dispersion Index, *DU*	.099	.102	.095	.093
	(4.43)	(3.60)	(3.62)	(3.57)
	.106	.105	.093	.092
	(4.31)	(3.44)	(3.15)	(3.15)
Guidepost Dummy, D_G	−.017	−.017	−.013	−.013
	(8.12)	(7.96)	(5.86)	(5.91)
	−.018	−.018	−.014	−.014
	(6.95)	(6.81)	(4.70)	(4.73)
Expected Price Increases: Consumer Price Index, $g_{C_{-1}}$	.136	.139	−.073	
	(2.15)	(2.03)	(.82)	
	.092	.089	−.050	
	(1.37)	(1.21)	(.58)	
Expected Price Increases: *Survey of Consumer Finances*, $EXPT_{-1}$		−.022	.381	.281
		(.14)	(2.04)	(2.00)
		.019	.348	.274
		(.10)	(1.59)	(1.56)
Unexpected Price Increases: *Survey of Consumer Finances*, $CATCH_{-1}$			.488	.405
			(3.39)	(3.97)
			.443	.388
			(2.73)	(2.97)
Standard Error/R^2	.00604/.83	.00609/.83	.00562/.86	.00560/.86
	.00796/.76	.00801/.76	.00760/.79	.00756/.79
Durbin-Watson	0.83	.83	.88	.88
	1.80	1.80	1.95	1.94

Source: See Gordon, *Brookings Papers*, 2:1972.

Notes: Estimated coefficients for two-quarter equations (with their absolute *t*-statistics in parentheses) appear in the top half of each row; estimates (and *t*-statistics) for one-quarter equations appear in the bottom half. The sample period for all equations is 1954:1–1970:4.

Definitions given for one-quarter equations. Variables in the two-quarter equations are all two-quarter averages of the corresponding variable in the one-quarter equations. The symbol g_x represents the quarterly rate of increase of the variable X at an annual rate, and $g_{x_{-\tau}}$ the same quarterly rate of increase lagged τ quarters.

g_w = the dependent variable: W is an hourly earnings index for production workers in the private nonfarm economy, adjusted for overtime and changes in interindustry mix. The index has further been adjusted to include fringe benefits. (See Gordon (1972).)

U^* = Weighted unemployment rate. (See Perry (1970).)

DU = Unemployment dispersion index. (See Perry (1970).)

D_G = Guidepost dummy. Set equal to 0.25 in 1962:1, 0.50 in 1962:2, 0.75 in 1962:3, 1.0 in 1962:4 to 1966:4, 0.75 in 1967:1, 0.50 in 1967:2, 0.25 in 1967:3, and 0.0 in other periods.

C = Consumer Price Index.

$EXPT$ = Annual rate of inflation expected, forthcoming year *SCF*. See Appendix.

$CATCH$ = Unexpected inflation during preceding year, at annual rate, *SCF*. See equation (5).

TABLE 2—ECKSTEIN-BRINNER WAGE EQUATION

Variable	Original Equation	*EXPT* Added	*EXPT* and *CATCH* Added	E-B Expectations Excluded
Inverse of Conventional	.0016	.002	.002	.001
Unemployment Rate,	(8.32)	(8.19)	(7.29)	(6.37)
$1/U$	.0018	.002	.002	.001
	(6.29)	(6.01)	(5.28)	(4.67)
Guidepost Dummy, D_G	−.0069	−.009	−.008	−.009
	(3.96)	(4.16)	(3.64)	(3.97)
	−.0076	−.009	−.008	−.010
	(2.86)	(3.01)	(2.58)	(3.01)
Expected Price Increases:	.244	.157	−.163	
Consumption Deflator,	(2.25)	(1.29)	(.624)	
g_d^*	.113	.031	−.299	
	(.71)	(.18)	(.90)	
Inflation Threshold, g_d^T	.821	.784	.779	
	(3.95)	(3.78)	(3.79)	
	.966	.914	.886	
	(3.16)	(2.96)	(2.87)	
Expected Price Increases:		.182	.534	.706
Survey of Consumer		(1.54)	(1.90)	(7.22)
Finances, $EXPT_{-1}$		.201	.573	.662
		(1.13)	(1.56)	(4.62)
Unexpected Price Increases:			.377	.492
Survey of Consumer			(1.38)	(4.31)
Finances, $CATCH_{-1}$			.413	.438
			(1.16)	(2.78)
Standard Error/R^2	.00584/.84	.00579/.85	.00574/.85	.00630/.82
	.00892/.70	.00890/.70	.00888/.71	.00932/.67
Durbin-Watson	1.27	1.30	1.36	1.14
	2.24	2.26	2.29	2.12

Source and general notes: See Table 1.

Variable Definitions:

g_W = Dependent variable. Same as Table 1.

U = Conventional unemployment rate.

D_G = Guidepost dummy. Same as Table 1.

d = Personal consumption deflator.

g_d^* = Expected inflation variable with imposed weights: $g_d^* = 0.4\, g_{d-1} + 0.3 g_{d-2} + 0.2 g_{d-3} + 0.1 g_{d-4}$. (See Gordon (1972).)

g_d^T = Threshold inflation variable $= 0.0$ if $[(d_{-1} - d_{-5})/d_{-5}] \leq 0.05$; otherwise $= \frac{1}{2}[d_{-1} - d_{-5}/d_{-5} - 0.05]$. (See Eckstein-Brinner (1972) and Gordon (1972).)

EXPT = Same as Table 1.

CATCH = Same as Table 1.

the explanatory power of *EXPT* and the given, indirect measure of expected inflation on the assumption that there is no catch-up effect. This test consisted of adding $EXPT_{-1}$ to the original equation.[17] The second test was a test of the joint hypothesis that *EXPT* is the better measure of expectations and that there is a catch-up effect. It consisted of adding $EXPT_{-1}$ and $CATCH_{-1}$ to the original equation. Since the two hypotheses just mentioned interact, it is quite possible that $EXPT_{-1}$ may fail the first test even though $EXPT_{-1}$ and $CATCH_{-1}$ together pass the second one. This in fact occurred in one case.

---

[17] The one-quarter lag allows for a response delay and parallels the lag specification in Gordon's comparative study.

The a priori ranges for values of the coefficients of $EXPT_{-1}$ and $CATCH_{-1}$ are both between 0 and 1, though there is no a priori restriction on which of the two should be larger. The results obtained are reported in columns 2, 3, and 4 of Tables 1, 2, and 3. Columns 2 and 3 report coefficient estimates when $EXPT_{-1}$ and both $EXPT_{-1}$ and $CATCH_{-1}$ were added to the original equation. Column 4 presents coefficients estimates when the original expected inflation variables were dropped and $EXPT_{-1}$ and $CATCH_{-1}$ took their place.

In Perry's equation, $EXPT_{-1}$ was not significant when entered alone (Table 1, col. 2).[18] However, both $EXPT_{-1}$ and $CATCH_{-1}$ were significant and Perry's rate of change of the consumer price index was not when the two *SCF* variables were entered together (Table 1, col. 3). The superiority of the latter equation to Perry's original equation clearly supports the two *SCF* variables.

When $EXPT_{-1}$ was entered alone in the Eckstein-Brinner equation, both g_d^*, the distributed lag measure, and $EXPT_{-1}$ were significant in the two-quarter estimate, and neither were in the one-quarter estimate (Table 2, col. 2). The threshold variable g_d^T remained significant in these and all estimates for the sample period extending through 1970:4.

Again, a clearer verdict was obtained when the two *SCF* measures were entered together (Table 2, col. 3). Both $EXPT_{-1}$ and $CATCH_{-1}$ were then individually significant in the two-quarter estimate and $EXPT_{-1}$ was significant and $CATCH_{-1}$ barely missed significance in the one-quarter estimate. By contrast, the Eckstein-Brinner distributed lag variable was insignificant in both estimates. This left the threshold variable g_d^T as the only significant Eckstein-Brinner variable. When additional evidence concerning instability of the coefficient of g_d^T is considered, the verdict that emerges is one favorable to the *SCF* variables.

Except for a brief period in 1957–58, the threshlod variable is only positive during the high inflation period of the late 1960's and early 1970's. Not surprisingly, its coefficient in the original Eckstein-Brinner equation was not significant for shorter sample periods tested (1954:1–1966:4, 1954:1–1968:4). In fact, when the equation in Table 2, column 3, was re-estimated for those sample periods, g_d^T was never significant and g_d^* was only significant in the two-quarter equations, whereas $EXPT_{-1}$ was significant in all equations and $CATCH_{-1}$ in the two-quarter equations. Moreover, in those shorter sample periods, all estimates of the equation with only *SCF* variables had lower standard errors than the corresponding estimates of the original Eckstein-Brinner equation. In short, when the evidence from all three sample periods is considered together, it supports the *SCF* variables rather than the Eckstein-Brinner variables.

The tests performed with Gordon's equation resulted in a standoff between the *SCF* variables and Gordon's expected inflation variable. When $EXPT_{-1}$ was entered alone, both it and Gordon's distributed lag measure, g_d^e, were significant in two-quarter and one-quarter estimates (Table 3, col. 2).[19] When $CATCH_{-1}$ was

---

[18] Unless otherwise indicated, an individual coefficient will be said to be significant if it passes a one-tailed t-test at the 10 percent level, and a group of coefficients will be said to be all significantly different from zero if they pass the appropriate F-test at the 10 percent level.

[19] Gordon in fact uses four variables to estimate the distributed lag in g_d^e. Therefore, the appropriate test of the hypothesis that not all of these coefficients are equal to zero is an F-test of the reduction of variance due to

(over to p. 178)

TABLE 3—GORDON WAGE EQUATION

Variable	Original Equation	*EXPT* Added	*EXPT* and *CATCH* Added	Gordon's Price Variables Excluded
Unemployment Dispersion Index, *DU*	.047	.036	.024	.037
	(3.05)	(2.09)	(1.32)	(2.21)
	.049	.030	.037	.040
	(2.18)	(1.20)	(1.40)	(1.76)
Disguised Unemployment Rate, U^D	−.605	−.704	−.693	−.808
	(5.24)	(5.31)	(5.30)	(6.09)
	−.626	−.786	−.760	−.735
	(3.43)	(3.88)	(3.70)	(3.89)
Unemployment Rate of Hours, U^H	−.238	−.325	−.420	−.223
	(1.26)	(1.66)	(2.08)	(1.42)
	−.366	−.455	−.379	−.213
	(1.39)	(1.72)	(1.35)	(1.05)
Expected Price Increases: Consumption Deflator, g^e	.608	.539	1.075	
	(5.16)	(4.15)	(3.02)	
	.580	.449	.134	
	(2.96)	(2.14)	(.30)	
Product Price, $(g_p - g_d)$	.739	.843	1.102	
	(3.89)	(4.19)	(4.24)	
	.768	.921	.750	
	(2.45)	(2.88)	(1.97)	
Expected Price Increases: *Survey of Consumer Finances*, $EXPT_{-1}$		.202	−.208	.514
		(1.47)	(.72)	(4.05)
		.361	.587	.513
		(1.71)	(1.68)	(2.82)
Unexpected Price Increases: *Survey of Consumer Finances*, $CATCH_{-1}$			−.492	.420
			(1.61)	(3.48)
			.311	.433
			(.82)	(2.59)
Standard Error/R^2	.00412/.93	.00408/.93	.00402/.93	.00504/.88
	.00736/.80	.00722/.81	.00724/.81	.00764/.76
Durbin-Watson	1.50	1.57	1.69	1.12
	2.61	2.63	2.56	2.23

Source and general notes: See Table 1.

Variable Definitions:

g_W = Dependent variable. Same as Table 1; *DU* = Same as Table 1.

U^D = Disguised unemployment rate; U^H = Unemployment rate of hours. (See Gordon (1972).)

d = Personal consumption deflator.

g_d^e = Distributed lag on the quarterly rate of change of the personal consumption deflator, with weights estimated as a polynomial distributed lag, starting in t−1, with individual coefficients constrained to lie along a fourth degree polynomial with level set equal to zero in the most distant period (t−12). The coefficients in the table are the sum of the estimated weights. The numbers in parentheses are the ratio of this sum to its standard error.

$(g_p - g_d)$ = Distributed lag on the difference between the quarterly rate of change of p, the private nonfarm deflator, and the quarterly rate of change of d, starting in t−1, with individual coefficients constrained to lie along a third degree polynomial and set equal to zero in the most distant period (t−7). The coefficients are sums of estimated weights, and each number in parentheses is the ratio of this sum to its standard error.

EXPT = Same as Table 1; *CATCH* = Same as Table 1.

added, the two-quarter estimate favored Gordon's variable while the one-quarter favored the *SCF* variables,[20] and judgment was again left in balance.[21]

In summary, if one views the tests performed on all the equations as a whole, the evidence supports both the inclusion of *EXPT* as an explanatory variable and the contention that there is a catch-up effect. This effect has interesting policy implications which can be best presented by decomposing *CATCH* into its parts and then rearranging the terms as follows:

$$\begin{aligned}(6)\quad \gamma\, EXPT_{-1} + \lambda\, CATCH_{-1} &= \lambda\, g^4_{d_{-1}} + (\gamma - \lambda) EXPT_{-1} \\ &\quad + \lambda\,(EXPT_{-1} - \overline{EXPT}_{-5})\end{aligned}$$

It is immediately clear that expected inflation is not the only feedback from prices to wages. Recent inflation, as represented by g^4_d, also influences wage increases. In fact, if $\gamma=\lambda$, a hypothesis which cannot be rejected in any of the estimates in column 4 of the three tables, the effect of expected inflation is only a transient one. Consequently, policy makers considering measures to reduce the rate of wage increase must consider the effects on unexpected as well as expected inflation. Pronouncements designed to bring down inflationary expectations—of which there have been many in the United States recently—may have little lasting effect if they do not also reduce actual inflation.

These tests also provide admittedly limited evidence concerning the stability of the tradeoff between inflation and unemployment. If all inflation were anticipated, *CATCH* would equal zero, and therefore, the existence of a tradeoff would depend solely on the magnitude of the coefficient of *EXPT*. In the estimates reported in column 4 of each table, this coefficient lies between .3 and .7 and is always significantly different from one. However, no allowance has been made in these estimates for the possibility that *EXPT* may be subject to measurement error, and that the estimates may consequently be biased downwards. Moreover, the sensitivity of wages to expected inflation under conditions of steady, anticipated inflation might well differ from their sensitivity in a period like that after the Korean War when even the expected rate of inflation was subject to considerable variation.

IV. Conclusion

The tests performed in Sections II and III demonstrate that the direct measures of expected and unexpected inflation developed here using responses to the *Survey of Consumer Finances* explain wage changes as well as or better than the indirect measures used in three recent studies. The significance of this result lies in its implication for our understanding of expectations. As the correlations in Section II indicate, the *SCF* measure of expected inflation differs greatly from the indirect measures of Perry, Eckstein and Brinner, and Gordon. If the *SCF* measure is correct, the other measures must be incorrect. But that observation constitutes a damning judgment, since the other measures represent the current state of knowledge regarding expectations of inflation.

---

those four variables jointly. The appropriate test of the hypothesis that the sum of the coefficients is positive is a one-tailed *t*-test based on the ratio of that sum to its standard error. In Table 3, col. 2, g_d^2 passes both the *F*-test and the *t*-test. (Only the *t*-statistic is reported in the table).

[20] In the two-quarter estimate, the *SCF* variables both have the wrong sign and Gordon's expected inflation variable passes both the *F*-test and the *t*-test. In the one-quarter estimate, however, $EXPT_{-1}$ is significant, and Gordon's variable passes the *F*-test, but not the *t*-test.

[21] Results for sample periods beginning in 1966:4 and 1968:4 were very similar, and do not alter this standoff decision.

Such a judgment leads one to seek a better understanding of expectations. Though that endeavor will be the subject of another study, hopefully this paper will have provided some of the data needed to extend the knowledge of expectations beyond the limited state to which it has been confined by the deficiencies of indirect measurement.

APPENDIX

The *Survey of Consumer Finances* (*SCF*), administered by the Survey Research Center, University of Michigan, is a survey of a representative sample of *U.S.* households, numbering around 3000, and has been conducted in January or February of every year since 1946. Supplemental surveys of about half the original size were added in 1952, and since 1959 the survey and its supplements have been administered roughly on a quarterly basis. The questions asked have dealt with both demographic and financial characteristics and with attitudes about economic conditions.

A question regarding the direction of price change similar to the one quoted in Section I has been included in almost every *SCF* since the inception of the survey in 1946. Raw data on the percentage of responses by category were obtained from reports on the *SCF* in several issues of the *Federal Reserve Bulletin* for 1946 through 1952, from *Consumer Expectations*, by George Katona and Eva Mueller for 1953–56, and subsequently from 1960 *Survey of Consumer Finances* and the succeeding annual publications of the Survey Research Center. Since 1968 the annual publications of the Center have not reported separately the percentage of respondents expecting prices to stay the same, go down, and the percentage uncertain. Those data were kindly supplied to us by Jay Schmiedeskamp, director, and Evelyn Hansmire of the Survey Research Center.

The wording of the question regarding price expectations has varied from time to time. Since 1959 the wording has been as reported in Section I. From 1951 to 1959 the question asked about "prices of household items and clothing," and from 1946 to 1951 about "prices of things in general." Overlapping periods were always available when the question changed form, and the overlap was used to link the different series.

Since 1959, the survey has usually been conducted four times a year. The timing of the surveys, however, has not been consistent, nor have survey dates necessarily matched the midpoints of the calendar quarters (February 15, May 15, August 15, and November 15). Prior to 1959, the survey was taken with varying frequency during each year. Consequently, whenever data were not available, or available at off times during the year, the series, which had been linked as described above, was interpolated linearly.

The question added to the survey in May 1966 asks for quantitative estimates from those respondents who expect prices to rise. The answers to this question were used to estimate the mean and standard deviation of expected price increases, *EXPT* and *SIG*. These answers have been coded in intervals (1–2 percent, 3–4 percent, 5 percent, 10 percent or more, don't know how much).

A mean of positive expected price increases was calculated by assigning the midpoint to each interval and 10 percent to the "10 percent or more" category. By assuming that the mean thus calculated reflected the average expectation of the "don't know how much" category, and by assigning a prediction of zero to the "fall" and "same" categories, a mean was obtained for all respondents who expressed a definite opinion in their answer to the first question. The standard deviation for this class of respondents was calculated in a consistent manner. The remainder of the sample, consisting of those who either did not answer the first question or answered that they were uncertain whether prices would rise, fall, or stay the same, was assumed to be distributed as the rest of the sample. Therefore the mean and standard deviation just described were taken as referring to the entire sample.

The Index of Consumer Sentiment was

obtained from *Survey of Consumer Finances*, 1967, pp. 247–48, and 1970, p. 207. This series was linearly interpolated in the same way as the price expectation series.

REFERENCES

J. A. Carlson and M. Parkin, "Inflation Expectations," mimeo, Purdue Univ. and Univ. of Manchester, May 1973.

——— and H. Ryder, "Quantitative Expectations From Qualitative Surveys: A Maximum-Likelihood Approach," mimeo, Purdue Univ. and Brown Univ., Oct. 1973.

G. de Menil, "Rationality in Popular Price Expectations," mimeo, Princeton Univ., Aug. 1973.

O. Eckstein and R. Brinner, "The Inflation Process in the U.S.," study for The Joint Economic Comm., 92d Cong., 2d Session 1972.

W. E. Gibson, "Interest Rates and Inflationary Expectations: New Evidence," *Amer. Econ. Rev.*, Dec. 1972, *62*, 854–65.

R. J. Gordon, "Inflation in Recession and Recovery," *Brookings Papers*, Washington 1971, *1*, 105–58.

———, "Wage Price Controls and the Shifting Phillips Curve," *Brookings Papers*, Washington 1972, *2*, 385–422.

D. Jaffee, "An Empirical Study of the Risk Structure of Interest Rates," Financial Res. Center, Princeton Univ., res. memo. 16, 1973.

T. Juster and P. Wachtel, (1972a) "Inflation and the Consumer," *Brookings Papers* Washington 1972, *1*, 71–114.

——— and ———, (1972b) "Anticipatory and Objective Models of Durable Goods Demand," *Amer. Econ. Rev.*, Sept. 1972, *62*, 564–79.

G. Katona and E. Mueller, *Consumer Expectations, 1953–1956*, Surv. Res. ser. no. 16, Inst. Soc. Res., Univ. Michigan, 1956.

E. Mueller, "Consumer Reactions to Inflation," *Quart. J. Econ.*, May 1959, *73*, 246–62.

J. F. Muth, "Rational Expectations and the Theory of Price Movements," *Econometrica*, July 1961, *29*, 315–35.

G. L. Perry, "Changing Labor Markets and Inflation," *Brookings Papers*, Washington 1970, *3*, 411–41.

D. H. Pyle, "Observed Price Expectations and Interest Rates," *Rev. Econ. Statist.*, Aug. 1972, *54*, 275–80.

H. Shufford, "Subjective Variables in Economic Analysis: A Study of Consumers' Expectation," unpublished doctoral dissertation, Yale Univ. 1970.

M. E. Sparrough and L. W. Bolton, "Calendar of Wage Increases and Negotiations for 1972," *Mon. Lab. Rev.*, Jan. 1972, *95*, 3–14.

L. D. Taylor, "Price Expectations and Households Demand for Financial Assets," mimeo, Univ. of Arizona and Univ. of Michigan, 1973.

H. Theil, *Economic Forecasts and Policy*, 2d ed., Amsterdam 1965.

S. J. Turnovsky, "Empirical Evidence on the Formation of Price Expectations," *J. Amer. Statist. Assn.*, Dec. 1970, *65*, 1441–54.

——— and M. L. Wachter, "A Test of the 'Expectations Hypothesis' Using Directly Observed Wage and Price Expectations," *Rev. Econ. Statist.*, Feb. 1972, *54*, 47–54.

C. T. Ward, "The 1966 Wage Calendar—Major Agreements and Reopenings," *Mon. Lab. Rev.*, Dec. 1965, *88*, 1426–37.

———, "Major Agreements, Expirations and Reopenings," *Mon. Lab. Rev.*, Dec. 1959, *82*, 1312–23.

Federal Reserve Bull., various issues, Washington, 1946–52.

Michigan Survey Research Center, *Survey of Consumer Finances*, annual vols., 1960–70. Ann Arbor.

Efficient and Optimal Utilization of Capital Services

By GUILLERMO A. CALVO*

In recent times there has been renewed interest in the factors determining capacity utilization in connection with the trade cycle.[1] However, little attention has been paid to the question of the *desirability* of different rates of capacity utilization. In the present paper I discuss this problem in the context of the neoclassical one-sector model when capital depreciates as a function of the intensity at which it is operated.

I. The Model

Let us call:

K_t = capital services per unit of time available at t (proportional to capital stock)

S_t = capital services per unit of time utilized at t

L_t = labor services per unit of time available at t

N_t = labor services per unit of time used at t

Y_t = output at t

C_t = consumption at t

δ_t = rate of depreciation per unit of available capital services per unit of time at t

We assume:

$$(1) \quad \frac{DL_t}{L_t} = n, \text{ exogenously given, } n > 0 \quad L_0 > 0$$

$$(2) \quad Y = F(S, N) \qquad \text{for all } S \geq 0,\ N \geq 0$$

(3)(a) F linear homogeneous, and continuous for all $S > 0$, $N > 0$

(b) $F(0, N) = 0$

(c) $F(S, 0) = 0$

(d) $\dfrac{\partial F}{\partial S} > 0$

(e) $\dfrac{\partial F}{\partial N} > 0 \qquad$ for all $S, N > 0$

(f) $\dfrac{\partial^2 F}{\partial S^2} < 0$

(g) $\dfrac{\partial^2 F}{\partial N^2} < 0$

(h) $\lim\limits_{S \to 0} \dfrac{\partial F}{\partial S} = \infty \qquad$ if $N > 0$

$$(4) \quad S_t = m_t K_t, \qquad 0 \leq m_t \leq 1$$
$$K_t \geq 0$$

$$(5) \quad 0 \leq N_t \leq L_t$$

$$(6) \quad DK_t = Y_t - C_t - \delta_t K_t, \qquad C_t \geq 0$$

Furthermore,

$$(7) \quad \delta = g(m) \qquad \text{for all } 0 \leq m < 1$$

where

(8)(a) $g(m)$ is twice-differentiable at every $0 \leq m < 1$

(b) $g'(m) \geq 0 \qquad$ for all $0 \leq m < 1$

(c) $g''(m) > 0 \qquad$ for all $0 \leq m < 1$

(d) $\lim\limits_{m \to 1} g(m) = \infty$

Restricting m to belong to the closed unit interval is equivalent to saying that capital services used are smaller than or equal to its availability and that they cannot be negative; m can be given at least two rather different economic interpretations:

---

* Assistant professor, Columbia University. The present work was completed while I was visiting professor at the Centro de Investigaciones para el Desarrollo (*C.I.D.*), Bogota, Colombia, during 1970.

[1] See Robert Lucas, and Paul Taubman and Maurice Wilkinson.

a) It can be a measure of the speed at which machines are operated per unit of time with constant labor services used per unit of time; or

b) it can measure the share of time per unit of time at which machines are operated, given a capital-labor services ratio. In other words, it may provide a model for the choice of the number of shifts, when the ratio of factor services is the same for every shift.

The latter is perhaps better understood if we define $\overline{N}(K, mK/N)$ as the flow of labor services required to operate capital at full capacity, i.e., when $S=K$, with capital-labor services ratio equal to mK/N. Clearly, then,

$$\frac{K}{\overline{N}\left(K, \frac{mK}{N}\right)} = \frac{mK}{N}$$

and, hence,

$$m\overline{N} = N$$

Therefore, recalling (3a), output when total capital services are K and the factor services ratio equals mK/N becomes

$$F\left(mK, m\overline{N}\left(K, \frac{mK}{N}\right)\right) = mF\left(K, \overline{N}\left(K, \frac{mK}{N}\right)\right)$$

where in view of our previous definitions, $F(K, \overline{N})$ is the full capacity output when the factor services ratio equals mK/N and the total supply of capital services is K. So m can be conceived as the share of the maximum output obtained when machines are operated m parts of a unit of time per unit of time with a given factor services ratio. In what follows, m will be referred to as the *utilization ratio*.

The rate of depreciation δ is by (8) a positive strictly convex function of m; positivity of δ is a natural assumption to make, whereas convexity is assumed for the sake of analytical convenience; the condition $\lim_{m\to 1} g(m)=\infty$ is innocuous and it will be used to guarantee that efficient m's are less than unity; however, there may be some justification for it if we interpret $m=1$ as the case where machines are operated at a maximum speed or without stopping for repairs (recall interpretations (a) and (b) above).

II. Efficient and Competitive Utilization Ratios

By (3) and (5) we can assure that it is efficient[2] to employ all the available labor services;[3] thus $N_t=L_t$ for every t.

Let us define

$k_t = \frac{K_t}{L_t}$ = ratio of available capital services to available labor services at time t

$$y_t = \frac{Y_t}{L_t}, \qquad c_t = \frac{C_t}{L_t}$$

So we can write, recalling (3),

$$y_t = f(m_t k_t) \tag{9}$$

where $f(x) = F(x, 1)$

Moreover, by (3),[4]

(10)(a) $f(x)$ is continuous for all $x \geq 0$

(b) $f(0) = 0$

(c) $f'(x) > 0$

(d) $f''(x) < 0$ for all $x > 0$

---

[2] As usual we call efficient any feasible per capita consumption path such that there is no other feasible per capita path yielding larger consumption (for a non-null set of points of the real line).

[3] Notice that this would not necessarily be the case if isoquants were L-shaped.

[4] Tjalling Koopmans, p. 236, has shown that if equations (3a–g) hold, there is some $\hat{x}$ such that $f(x) \leq nx$ for all $x \geq \hat{x}$. The latter together with (10c and d) readily implies (10f).

(e) $$\lim_{x\to 0} f'(x) = \infty$$

(f) $$\lim_{x\to 0} [f(x) - nx] = -\infty$$

Due to the fact that we are in an essentially one-good economy, static efficiency suffices to determine the value of m_t. Clearly, we will choose m so as to

$$\max_{1\geq m\geq 0} [f(mk_t) - g(m)k_t] \tag{11}$$

By (8) and (10) there exists a unique interior solution. It satisfies the following first-order necessary condition when $k_t > 0$,

$$f'(mk_t) = g'(m) \tag{12}$$

In other words, efficient utilization of capital at a given point in time t calls for equating the marginal productivity, $f'(mk_t)$, to the marginal cost of capital services at t, $g'(m)$. Figure 1 shows how the efficient level of m—indicated by $m^*(k_t)$—is determined. Notice that since $f'' < 0$, f' will shift to the left in Figure 1 as k_t is increased; hence,

$$\frac{dm^*(k)}{dk} < 0 \tag{13}$$

Thus, the efficient utilization ratio decreases as the ratio of the available capital to labor services k is increased. To put it differently, the richer is this idealized economy in terms of available capital services per capita, the smaller will be the (efficient) share of total available capital services utilized.

By (8) and (12) it readily follows

$$\lim_{k\to 0} m^*(k) = 1 \tag{14}$$

On the other hand, in a competitive situation the (instantaneous) real rate of interest r should in equilibrium be equal to the net marginal productivity of the capital stock, i.e.,

$$f'(m_t k_t)m_t - g(m_t) = r_t \tag{15}$$

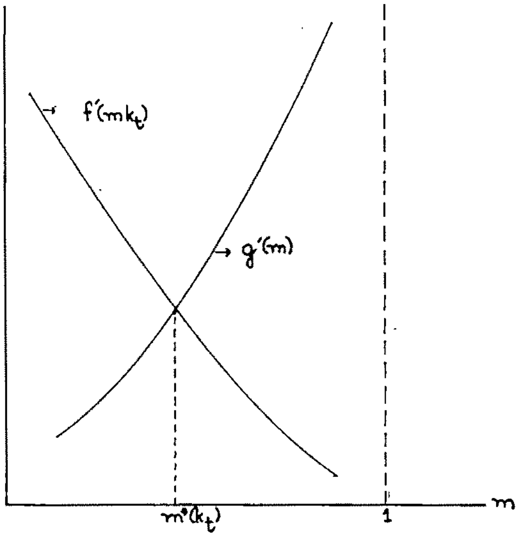


FIGURE 1

Moreover, it is clear that profit maximization implies $m_t = m^*(k)$. Hence, by (12) and (14), we get

$$g'(m_t)m_t - g(m_t) = r_t \tag{16}$$

By implicit differentiation of (16) we obtain

$$\frac{dm_t}{dr_t} = \frac{1}{g''(m_t)m_t} > 0 \tag{17}$$

i.e., the larger is the rate of interest the larger will also be the utilization ratio.

An interesting implication of our assumptions is that by (16), the utilization ratio can be determined as soon as r_t is known, and that calculations can be performed by using function g exclusively.

Finally, differentiating (12) totally with respect to k yields

$$f''(mk)\frac{d(mk)}{dk} = g''(m)\frac{dm^*(k)}{dk} \tag{18}$$

and since $dm^*(k)/dk < 0$, we get

$$\frac{d(mk)}{dk} > 0 \tag{19}$$

which means that the ratio of capital services *used* to the labor services *used*, mk, is positively correlated to the corresponding ratio of *available* factor services k. Notice that this is obtained despite the fact that the efficient utilization ratio $m^*(k)$ is negatively correlated with k (recall (13)).

III. Optimal Utilization Ratios

In view of (1)–(7), the notation introduced in Section II, and assuming efficiency, one can write without loss of generality,

$$(20)\quad \begin{aligned} Dk_t &= f(m_t k_t) - [g(m_t) + n]k_t - c_t \\ &= h(k_t) - c_t \end{aligned}$$

where

$$(21)\quad h(k) \equiv f(m^*(k)k) - [g(m^*(k)) + n]k$$

Clearly, $f(k) - nk \geq f(mk) - [\delta + n]k$, and hence, by (10f),

$$(22)\qquad \lim_{k\to\infty} h(k) = -\infty$$

On the other hand, recalling (12), for all $0<k$ we have

$$(23)\quad \begin{aligned} h'(k) &= f'(m^*(k)k)[m^{*\prime}(k)k + m^*(k)] \\ &\quad - g'(m^*(k))m^{*\prime}(k)k - g(m^*(k)) - n \\ &= f'(m^*(k)k)m^*(k) - g(m^*(k)) - n \\ &= g'(m^*(k))m^*(k) - g(m^*(k)) - n \end{aligned}$$

hence, by (13) and (23), for all $k>0$,

$$(24)\quad h''(k) = g''(m^*(k))m^*(k)m^{*\prime}(k) < 0$$

For any given $0 \leq m < 1$, (10) can be used to show that there is some $\bar{k} > 0$ such that for all $0 < k < \bar{k}$

$$(25)\quad f(mk) - [g(m) + n]k > 0$$

while

$$(26)\quad f(mk) - [g(m) + n]k = 0 \quad \text{for } k = 0$$

Hence, since $m^*(k)$ solves (11), then (22), (24), and (25) imply the existence of some $\bar{k} > 0$ such that

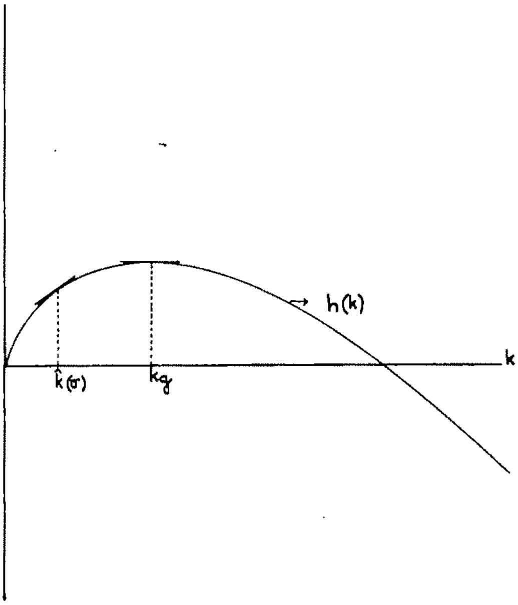


FIGURE 2

$$(27)\quad h(k) > 0 \quad \text{if and only if} \quad 0 < k < \bar{k}$$

Furthermore, by (8b) and (26),

$$(28)\qquad h(0) = 0$$

Figure 2 depicts a function satisfying (24), (27), and (28).

We are now ready to prove the first central result of this section.

PROPOSITION 1: *There exists one and only one golden rule capital-labor ratio, k_g. The corresponding unique golden rule utilization ratio m_g satisfies*

$$(29)\qquad g'(m)m - g(m) = n$$

PROOF:

From (20), golden rule consumption per capita c_g satisfies

$$(30)\qquad c_g = \max_{k\geq 0} h(k) = h(k_g)$$

By (24), (27), and (28), k exists and is unique. Also $h(k_g) > 0$ (see Figure 2). Clearly, by (23) and (30),

$$0 = h'(k_0) = g'(m^*(k_0))m^*(k_0) - g(m^*(k_0)) - n \tag{31}$$

By (8), $g'(m)m-g(m)$ is a strictly increasing function of m, $0 \leq m < 1$; hence, there is a unique m for which (31) holds. Existence of m_g follows from the fact that $m_g = m^*(k_g)$ for which existence has been proved above.

Remark: Equation (16) above and Proposition 1 imply that, as usual, the (instantaneous) real rate of interest in the golden rule path equals the rate of growth of labor. In fact, given the latter and the depreciation function, one can use (16) to determine m_g. The golden rule value of the utilization ratio becomes, under the present assumptions, completely independent of the production function.

The present model can easily be suited to study optimal utilization ratios in an optimal growth plan. Consider the following problem:

$$\max \int_0^\infty u(c_t)e^{-\sigma t}dt \qquad \sigma > 0 \tag{32}$$

subject to $Dk_t = h(k_t) - c_t$

$$k_t \geq 0, \quad c_t \geq 0 \qquad \text{for every t}$$

$$k_0 \text{ given}$$

where u is assumed to be strictly concave and increasing, and to satisfy $\lim_{c \to 0} u(c) = -\infty$. Functions associated with the solution of (32) will be called optimal. Let us define $\hat{k}(\sigma)$ as the value of k such that (see Figure 2)

$$h'(\hat{k}) = \sigma \tag{33}$$

$\hat{k}(\sigma)$ exists and it is unique. Existence is shown in the Appendix; uniqueness, on the other hand, follows from (24).

PROPOSITION 2: *For any initial capital-labor ratio* $\bar{k} > k_0 > 0$,[5] *there exists a unique optimal path of* k*'s;* k_t *converges monotonically to* $\hat{k}(\sigma)$. *The associated* $m_t = m^*(k_t)$ *converges monotonically to* $m^*(\hat{k}(\sigma))$.

PROOF:

Existence and monotonic convergence of k will not be proved here since it involves a trivial application of results obtained by Koopmans, for example.

Monotonicity and convergence of m_t follows directly from the properties of the optimal path of k's and (13).

Remark: Since by (13) there is an inverse relationship between m and k, we have that if in an optimal plan $k_0 < \hat{k}(\sigma)$ $[k_0 > \hat{k}(\sigma)]$, then $m_0 > m^*(\hat{k}(\sigma))$ $[m_0 < m^*(\hat{k}(\sigma))]$; therefore optimal k_t increases (decreases) and optimal m_t decreases (increases) over time. By (23) and (33) we have

$$\sigma = h'(\hat{k}(\sigma)) = g'(m^*(\hat{k}(\sigma))m^*(\hat{k}(\sigma)) - g(m^*(\hat{k}(\sigma))) - n \tag{34}$$

But, by a similar argument as in the previous Remark, one can show that there is one and only one m for which (34) holds true. The asymptotic value of m is again wholly independent of the production function.

In view of (16) and the results of this section it would be interesting to find out about the family of functions $z(m)$ satisfying (8) such that

$$z'(m)m - z(m) = g'(m)m - g(m) \qquad \text{for every } 0 \leq m < 1 \tag{35}$$

It can easily be shown that every function belonging to that family can be expressed as

$$z(m) = g(m) + bm, \quad \text{for some } b \text{ such that } -bm \leq g(m), \qquad \text{for all } 0 \leq m < 1 \tag{36}$$

Therefore, golden rule and optimal utilization ratios are not going to be affected by changing the depreciation function from g to z (satisfying (8)) as long as $z'(m)$

---

[5] Recall the definition of $\bar{k}$ in (27).

$-g'(m)=b=$ constant for all $0\leq m<1$, and $z(0)=g(0)$.

APPENDIX

LEMMA: *Let* $\sigma>0$; *then there is some* $k>0$ *such that* $h'(k)=\sigma$.

PROOF:

Taking into account (22), (24), and (27), it is sufficient to show

$$\text{(A1)} \qquad \lim_{k\to 0} h'(k) = \infty$$

By (14) and (23), the latter is equivalent to

$$\text{(A2)} \qquad \lim_{m\to 1} [g'(m)m - g(m)] = \infty$$

We will show that (A2) holds given that g satisfies (8). First notice that

$$\text{(A3)} \qquad \frac{d}{dm}[g'(m)m - g(m)] = g''(m)m > 0$$

Suppose now that (A2) is false. Then by (A3),

$$\text{(A4)} \qquad \lim_{m\to 1} [g'(m)m - g(m)] = B, \qquad B \text{ a real number}$$

Moreover, by (A3)

$$\text{(A5)} \qquad B > -g(0)$$

Due to (8c and d) there is some $\bar{m}<1$ such that

$$\text{(A6)(a)} \qquad g(\bar{m}) > 0$$

$$\text{(b)} \quad g(m) \text{ is increasing for all } 1 \geq m \geq \bar{m}$$

Thus (A4) is equivalent to

$$\text{(A7)} \qquad \lim_{m\to 1} \left[\frac{g'(m)m}{g(m)} - 1\right] g(m) = B$$

The latter implies, recalling (8d), (A5), and (A6),

$$\text{(A8)} \qquad \lim_{m\to 1} \frac{g'(m)}{g(m)} = 1$$

or equivalently,

$$\text{(A8')} \qquad \lim_{m\to 1} \frac{d\, ln\, g(m)}{dm} = 1$$

Hence, for every $\epsilon>0$, there is some $0<\hat{m}(\epsilon)<1$, such that

$$\text{(A9)} \qquad \frac{d\, ln\, g(m)}{dm} < 1 + \epsilon, \qquad \text{for all } 1 \geq m > \hat{m}(\epsilon)$$

Therefore,

$$\text{(A10)} \qquad ln\, g(m) < ln\, g(\hat{m}(\epsilon)) + (1+\epsilon)(m - \hat{m}(\epsilon)) \qquad \text{for all } 1 \geq m > \hat{m}(\epsilon)$$

which is equivalent to

$$\text{(A11)} \qquad g(m) < g(\hat{m}(\epsilon)) \exp\left[(1+\epsilon)(m-\hat{m}(\epsilon))\right] \qquad \text{for all } 1 \geq m > \hat{m}(\epsilon)$$

Hence,

$$\text{(A12)} \qquad g(1) < g(\hat{m}(\epsilon)) \exp\left[(1+\epsilon)(1-\hat{m}(\epsilon))\right] < \infty$$

contradicting (8d). Therefore (A2) holds and the Lemma follows.

REFERENCES

T. C. Koopmans, "On the Concept of Optimal Economic Growth," in *The Econometric Approach to Development Planning*, Amsterdam 1965.

R. E. Lucas, Jr., "Capacity, Overtime and Empirical Production Functions," *Amer. Econ. Rev. Proc.*, May 1970, *60*, 23–27.

P. Taubman and M. Wilkinson, "User Cost, Capital Utilization and Investment Theory," *Int. Econ. Rev.*, June 1970, *11*, 209–15.

International Trade and Economic Growth: A Diagrammatic Analysis

By ROBERT E. BALDWIN*

As others have noted (see articles by Harry Johnson (1971, 1972) and Jaroslav Vanek), the highly mathematical nature of most of the literature integrating growth and trade theory serves to restrict considerably the size of the audience to which explanations of this important subject can be directed.[1] This paper attempts to widen this audience by framing the analysis of the topic in simple geometric terms that utilize the standard two-country, two-sector (commodity), two-factor model of international trade.

Assume that the two commodities produced by the two countries A and B are a capital good K and a consumption good C. The capital good together with a growing labor force L are the two factors of production.[2] As is typical for this model, also suppose that the capital good cannot be traded once it is used as a productive factor and that this good is produced in a uniformly labor-intensive fashion compared to the consumption good.[3] The production function of each good is identical in both countries and subject to constant returns to scale.[4] It is further assumed that the growth rate of the labor force is the same in both countries and that each economy saves a constant but different proportion of its income.[5]

There are two requirements for steady-state equilibrium in such a model: 1) The desired rate of capital accumulation in both countries must equal the exogenously determined rate of growth of the labor force, and 2) The value of exports must equal the value of imports for each of the two countries.[6] As the first step in the analysis, it will be shown that over some range of capital-labor (K/L) endowment ratios for each country, there is for each of these endowment ratios some relative price of the capital good in terms of the consumption good (P_k) that will equalize the desired rate of capital accumulation in each country and the common rate of labor force growth. After finding the set of P_k and K/L ratios for each country satisfying this equilibrium condition, the next step will be to select the subset of P_k at which the two countries wish to trade in opposite directions, and determine for each of these P_k the relative size of the countries as measured by the ratio of their labor supplies that would be needed to equalize the value of exports and imports in each country. Given some actual initial ratio of the size of the two labor forces—a ratio that remains constant throughout the growth process, since the rates of labor force growth are equal in the two countries—it will then be possible to read off from this schedule of

---

* Professor of economics, University of Wisconsin, Madison.

[1] The exposition of these authors deals mainly with the case of a small country facing fixed terms of trade whereas the analysis here focuses on the more general case where there are two countries, each of which can affect the terms of trade.

[2] For simplicity, depreciation of the capital good is ignored. Some initial supply of K and L in each country is also assumed.

[3] This assumption concerning factor intensity is made to insure a unique, stable steady-state equilibrium.

[4] Such other standard assumptions of the simple Heckscher-Ohlin-Samuelson model as diminishing marginal productivity and perfect competition are also made.

[5] In each economy, all individuals have the same savings ratio and possess the same quantity of capital.

[6] When the *desired* rate of capital accumulation in a country equals the labor force growth rate, the country's supply of the capital good will generally not equal its demand for this good. However, the trading condition required for equilibrium insures that one country's excess demand for K is matched by the other country's excess supply. Thus, the two requirements can be combined into the more familiar condition that the *actual* rate of capital accumulation equal the common growth rate of the labor force.

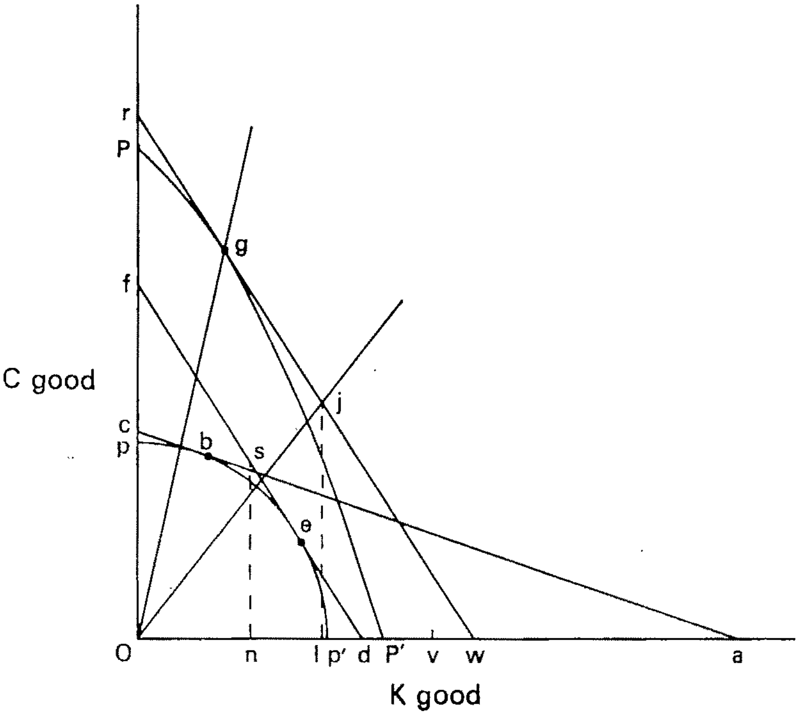


FIGURE 1

hypothetical relative sizes of the two countries that equalize trade at different P_k, the particular P_k and related L/K ratios that satisfy both equilibrium conditions.

I. Determining Possible Equilibrium Capital-Labor Ratios

To begin with the first step, the manner in which the desired rate of accumulation varies with changes in the relative price of the capital good is depicted in Figure 1. The curve pp' represents country A's per worker production possibilities for the two goods given a particular stock of capital per member of the labor force. At the relative price of the capital good indicated by the slope of the line abc, the value of output per worker measured in terms of the capital good would be Oa. If the saving ratio is $\frac{1}{2}$, the amount of capital per worker that the economy would desire to add to its capital stock would be Ov, i.e., $\frac{1}{2}$ of Oa. On the other hand, if P_k rises to the slope indicated by def, desired investment would decline to On, i.e., $\frac{1}{2}$ of Od. Thus, given a particular K/L endowment ratio, the lower the price of the capital good, the higher will be the desired rate of capital accumulation.[7]

This relationship between P_k and the rate of capital accumulation, given a fixed K/L endowment ratio, for example, K_1/L_1, is depicted in Figure 2. As can be seen from a consideration of Figure 1, for all prices of the capital good at which the country is completely specialized on this commodity, the rate of capital accumulation becomes fixed. The P_k, accumulation-rate curve, K_1/L_1, in Figure 2 thus becomes vertical. On the other hand, the rate of accumulation continues to rise as P_k falls below the price at which complete specialization on the consumption good occurs, i.e., below the price at which the curve cc' intersects the P_k, accumulation-rate curve, K_1/L_1.

Relating P_k and the rate of capital ac-

---

[7] The desired accumulation rate per worker is, of course, also the desired accumulation rate for the economy as a whole.

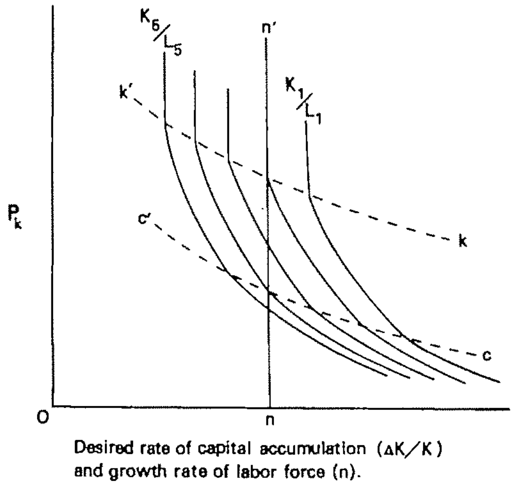


FIGURE 2

cumulation for a K/L ratio higher than K_1/L_1 results in a shift in the P_k, accumulation-rate curve to the left. This can be seen in the following way. Suppose the total capital endowment available to a given number of workers who produce both the capital and consumption good is increased, while the relative prices of the two goods remain fixed. As is well known from the analysis of this problem in terms of the Bowley-Edgeworth box diagram, the output of the two goods will be adjusted so that the marginal productivities of capital and labor remain unchanged. This means that the income accruing to capital increases in the same proportion as the increase in the capital stock, whereas the income earned by the fixed labor supply remains unchanged. The sum of the income earned by capital and labor, therefore, increases a smaller percentage than the increase in the capital stock. The volume of desired savings and investment, which is proportional to the level of total income, will consequently not rise as much relatively as the increase in K/L (which equals the proportionate increase in the capital stock), and the rate of accumulation will decline at any P_k.[8] As depicted by the curves kk' and cc' in Figure 2, the capital-good prices at which specialization on the capital good (shown by kk') and on the consumption good (shown by cc') occur, increase as K/L rises. This follows from the nature of the Rybczynski effect

---

[8] If only one good is produced before and after the increase in K, the fact that total income rises relatively less than the relative increase in K follows directly from the assumption that the production functions exhibit diminishing marginal productivity.

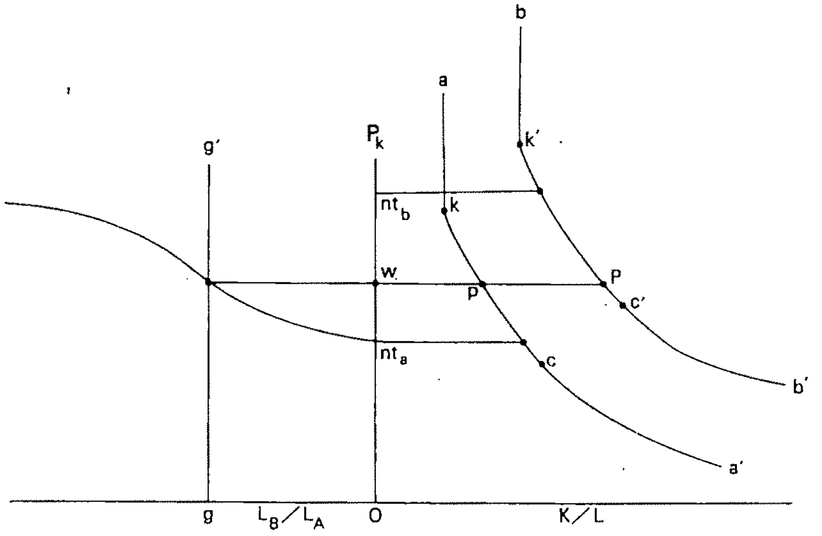


FIGURE 3

when the capital good is labor intensive compared to the consumption good.[9]

The distance *On* in Figure 2 indicates the given rate of labor force growth. Therefore, by drawing a vertical line *nn'* at *On*, it is possible to read off all the P_k and K/L ratios which satisfy the condition that the desired rate of capital accumulation equal the rate of labor force growth. The curve *aa'* in Figure 3 depicts the set of such points for country *A*. Since the savings ratio in country *B* is assumed to be greater than in *A*, the curves in Figure 2 relating P_k to the rates of capital accumulation for various K/L ratios are further to the right for country *B* than for country *A*. This in turn implies that for country *B*, the set of points indicating all possible P_k and K/L ratios satisfying the equilibrium condition that the desired rate of capital accumulation equal the rate of labor force growth behaves like the curve *bb'* in Figure 3. The points where each curve becomes vertical (*k* and *k'*) indicate the P_k and K/L ratios at which specialization on the capital good occurs, while *c* and *c'* depict where specialization on the consumption good takes place for the two countries.

II. Introducing the Market-Clearing Condition

As previously mentioned, besides the equilibrium requirement that the desired accumulation rate in each country equal the common labor force growth rate, the excess domestic demand of one country for a particular commodity must equal the excess domestic supply of that commodity by the other country. Over some range of the P_k and K/L ratios that satisfy the first equilibrium condition, the two countries will wish to trade in opposite directions, i.e., a commodity that one country wishes to export the other country wishes to import. For each of these P_k there is some hypothetical relative size of the two countries, as measured by the ratio of their labor (or capital) supplies, that will satisfy the equilibrium requirement relating to trade.

Consider, for example, the P_k level *Ow* in Figure 3 and the corresponding K/L ratios, *wp* and *wP* for countries *A* and *B*, respectively. Suppose that the curves *pp'* and *PP'* in Figure 1 represent the per worker production-possibility curves of countries *A* and *B*, respectively, that are associated with these K/L ratios. Furthermore, let the capital-good price *Ow* be indicated in this figure by the common slope of the lines *def* and *wgr*. With this price ratio, country *A*'s per worker production point will be *e* and—assuming a savings ratio of $\frac{1}{2}$—each individual will wish to trade by exporting the capital good and importing the consumption good until the point *s* is reached, since $\frac{1}{2}$ of *Od* equals *On*. In country *B*, where the per worker savings ratio is assumed to be greater than $\frac{1}{2}$, the per worker production point will be at *g*, and each individual will desire to trade along the budget line *wgr* until the point *j* is reached (*Ol* is more than $\frac{1}{2}$ of *Ow*). Since the trading distance *gj* is longer (actually twice as long) than *es*, the two countries cannot be of equal size in terms of their labor endowments and still achieve a trading equilibrium at the given P_k. However, since at the given P_k each worker in country *A* wishes to export only one-half the volume of *K* that each worker in country *B* wishes to import, if there are twice as many workers in country *A* as in country *B* (each with the capital endowment of the country *A* worker whose production-possibility curve is depicted in Figure 1), country *A*'s total exports of *K* would equal the total volume of country *B*'s imports of this good. This ratio of the size of the labor force in country *B* (L_B) to the size in country *A* (L_A)—$L_B/L_A = \frac{1}{2}$ in this case—that is needed to equate exports and imports at the given P_k and K/L ratios is plotted in the left quadrant of Figure 3 at the given P_k.

Next select a lower P_k than *Ow* and consider the magnitude of the L_B/L_A ratio required for trading equilibrium at this lower P_k and its associated K/L ratios in the two countries. The right side of Figure 3 indicates that the growth-equilibrium K/L

---

[9] Increasing the volume of capital per worker shifts the per worker production-possibility curve outward in such a manner that the slope of the curve increases at its intersection points with both the capital-good and the consumption-good axes.

ratios for the two countries are larger than at the P_k ratio of Ow. It is convenient to think of the change in P_k and K/L ratios as taking place in two steps: first, by an increase in the K/L ratios with P_k remaining constant, and then by a decrease in P_k with the K/L ratios being held at their higher levels. Both of these changes act in both countries to increase per worker demand for the capital good and decrease per worker supply. For example, referring again to Figure 1, suppose that the ratio of capital per worker is increased from the ratio that gives the per worker production-possibility curve pp', while P_k is held constant at the level indicated by the slope of def. From the Rybczynski theorem we know that if the capital good is labor intensive, the per worker production-possibility curve will shift to positions such as PP', where at the previous P_k (equals the slope of wgr) the output of the capital good declines as compared to its previous level. Since income per worker measured in terms of the capital good and at the fixed P_k increases (from Od to Ow), the demand for the capital good increases. Similarly, if the price of the capital good in terms of the consumption good is decreased while the K/L ratio is kept fixed at its new higher level, production of the capital good declines further and income, and thus investment demand, increases still more.

Since country A exports the capital good in return for the consumption commodity whereas country B does the opposite, the effect of lowering P_k and increasing the K/L ratio in both countries will be to reduce A's per worker excess supply of the capital good and increase B's per worker excess demand for this good. Therefore, for the relative size of the two countries that produced trading equilibrium with the previous higher P_k, the demand for the capital good on the part of country B will exceed the supply from country A. To achieve trading equilibrium at the lower P_k and higher K/L ratios, it is necessary to decrease the size of country B relative to country A, i.e., reduce L_B/L_A. This will have the effect of bringing about trading equilibrium by increasing the supply of the capital good relative to the demand for this commodity. It follows, therefore, that the shape of the P_k, L_B/L_A equilibrium curve in Figure 3 will be such that as P_k decreases, L_B/L_A also decreases.

The P_k, L_B/L_A curve will range between the two P_k ratios that define the points where both countries will wish to export (or import) the same commodity, i.e., the curves will range between the no-trade points of the two countries. For country A, as P_k continues to decrease below Ow and the per worker K/L ratio increases beyond wp, excess supply will finally decline to zero at a point such as Ont_a, and the associated K/L ratio will be country A's steady-state equilibrium position in the absence of international trade. Since country B's K/L ratio at this P_k as well as its savings propensity is higher than A's, B will be a net demander of the capital good at the P_k that represents A's no-trade equilibrium. At P_k ratios below Ont_a, country B's excess demand for the capital good will be even greater, whereas country A will also wish to import the capital good. To find country B's no-trade equilibrium, P_k must be increased and the K/L ratio decreased. These changes will decrease per worker demand for the capital good and increase its supply. Thus, a P_k such as Ont_b will be where country B ceases to be a net demander of the capital good.[10]

---

[10] Unlike the situation depicted in Figure 3, the autarchy points Ont_a and Ont_b could lie in the P_k range where both countries produced both commodities. The equilibrium K/L ratio for country B at its no-trade equilibrium point must, however, be greater than the K/L ratio at A's no-trade equilibrium point. Consider, for example, whether the two equilibrium K/L ratios could be the same. With the same K/L ratio the total demand for K would be the same in both countries, since the rate of accumulation is the same at all points on both P_k, K/L curves. However, since the no-trade P_k for country B must be higher than for A, country B's production of K would be greater than for country A with the same K/L ratio. Thus if excess demand and supply were zero for country A, i.e., if it were at its no-trade equilibrium point, excess supply would have to be positive in country B. With a P_k higher than prevails in country A, lowering the K/L ratio for country B merely increases its excess supply. Thus a higher K/L is necessary in country B than in A in order to decrease its production of K and increase its demand for K at a given P_k and reduce excess supply to zero.

At points above Ont_b, both A and B will wish to export the capital good.

Consider the shape of the P_k, L_B/L_A curve as it approaches the two no-trade P_k ratios. As the P_k ratio Ont_b is approached from below, country A's per worker excess supply of the capital good increases and country B's per worker excess demand for this good decreases. However, by increasing the size of country B relative to country A, it will always be possible to make B's excess demand for the capital good sufficiently large to equal the excess supply of this good from A. The closer the capital-good price is to Ont_b, the higher the L_B/L_A ratio must become in order to make demand equal supply. Therefore, the equilibrium L_B/L_A ratio approaches infinity as Ont_b is approached. When Ont_b is reached, country A's excess supply of the capital good is negligible compared to B's consumption of this good, and country A in effect faces an infinitely elastic demand curve for the capital good. Conversely, at the P_k ratio Ont_a, country B is negligible in size compared to country A and in effect faces an infinitely elastic supply curve of the capital good.

III. Equilibrium Possibilities

The last step in determining the steady-state equilibrium P_k and K/L ratios for the two countries is to mark the actual initial ratio of the labor supplies in the two countries on the L_B/L_A axis in Figure 3. Since the labor supplies grow at the same rate in both countries, this ratio never changes for the two countries. If this ratio is Og, the P_k and the K/L ratio that satisfies both the trade and accumulation equilibrium conditions for steady-state growth are Ow for the capital-good price and wp and wP for the capital-labor ratios for countries A and B, respectively. Both of the countries are incompletely specialized when Og is the ratio of labor supplies in the two countries.[11] However, if initial relative labor supplies are such that the equilibrium P_k ratio is only slightly above Ont_a, country A will produce both goods but country B will be completely specialized on the consumption good. On the other hand, if the L_B/L_A ratio is sufficiently high to give an equilibrium P_k level only slightly below Ont_b, country A will become completely specialized on the capital good whereas country B will produce both goods. There are four additional possible patterns of specialization.[12] If the savings ratios for the two countries are simply reversed, then country B's P_k, K/L curve will lie to the left of country A's.[13] This makes it possible for country A to specialize in the production of the consumption good while country B produces both goods (exporting the capital good) and also for country B to specialize in the production of the capital good whereas country A produces both commodities. The last two possible cases arise if the savings ratios of the two countries differ so much that the P_k level at the point where one of the countries becomes completely specialized on the consumption good is higher than the P_k level at which complete specialization on the capital good by the other country begins. If country B has the higher savings ratio, this country will specialize on the consumption good, while country A will specialize on the capital good. Reversing the the positions of the two countries in these circumstances gives the other possible specialization case.

In all of these seven possibilities the country with the smaller savings ratio (and thus with the P_k, K/L curve further to the left in Figure 3) will have a lower P_k prior to trade and thus will increase is production of the capital good after trade is opened up and export this commodity. In a sense, the country with the lower savings ratio has a demand-induced comparative advantage in the labor-intensive good relative to the capital-intensive commodity.[14] The other country increases the production of the consumption good and exports this good. Since the capital good is labor intensive compared to the consumption good, these shifts in production—in accordance with the Stolper-Samuelson theorem—increase the wage rate and decrease the rental return to capital in the capital-exporting country as compared with its no-trade position. Factor returns in the other country move in the opposite direction as trade is opened up.

---

[11] Since both countries are incompletely specialized, the capital-labor ratio employed in producing each good will be the same in both countries, and the wage-rental ratio will be identical for the two countries.

[12] See Murray Kemp, pp. 222–24.

[13] Since the x-axis in the left-hand part of Figure 3 is the ratio of the labor supply in the country with the higher savings propensity to the labor supply in the country with the lower savings propensity, consistency requires this axis to be labelled the ratio of L_A to L_B.

Consider P_k somewhat above and below the point where the P_k, L_B/L_A curve intersects the vertical line gg', which indicates the actual L_B/L_A ratio for the two countries. Since at P_k above this intersection point the actual L_B/L_A ratio is less than the L_B/L_A ratio that would fulfill the trading equilibrium condition, country B's excess demand for the capital good will be less than country A's excess supply. The price of the capital good would therefore tend to decline and the K/L ratios in both countries to rise in response to the fall in the price of the capital good. By similar reasoning, P_k ratios which are below the equilibrium level result in country B's excess demand for the capital good being greater than country A's excess supply. This relationship in turn tends to increase the price of the capital good and reduce the K/L ratios in the two countries. Hiroshi Atsumi has shown that when the trading-equilibrium position for the commodity offers associated with all the possible steady-state P_k, K/L ratios (as shown by the aa' and bb' curves in Figure 3) is stable, the growth paths of the two countries will converge to the equilibrium L_B/L_A, P_k, and K/L ratios indicated at the intersection point of the P_k, L_B/L_A curve and the gg' line.

The effects of such parametric changes as an increase in one country's savings ratio, an increase in the common rate of labor force growth (or even the introduction of unequal rates of labor force growth), and the imposition by one of the countries of a tariff on either the capital or consumption good can easily be analyzed with the aid of Figures 2 and 3. By causing the P_k, K/L curves in Figure 3 to shift, these changes lead to shifts in the P_k, L_B/L_A curve in that diagram and thus to new equilibria for P_k and the K/L ratios.

REFERENCES

H. Atsumi, "The Long-Run Offer Function and a Dynamic Theory of International Trade," *J. Int. Econ.*, Aug. 1971, *1*, 267–99.

H. Johnson, "Trade and Growth: A Geometrical Exposition," *J. Int. Econ.*, Feb. 1971, *1*, 83–101.

———, "Trade and Growth: A Correction," *J. Int. Econ.*, Feb. 1972, *2*, 87–88.

M. Kemp, *The Pure Theory of International Trade and Investment*, Englewood Cliffs 1969.

J. Vanek, "Economic Growth and International Trade," *Quart. J. Econ.*, Aug. 1971, *85*, 377–90.

---

[14] If the savings ratios are the same in both countries, the steady-state position will be characterized by identical K/L ratios for the two countries, and no trade. In terms of Figure 3 there would be only one P_k, K/L curve in the right-hand quadrant and the P_L, L_B/L_A curve on the left-hand side would be a horizontal line at the P_k level indicating the no-trade position on the P_k, K/L curve.

Lancaster's New Approach to Consumer Demand and Its Limitations

By REUVEN HENDLER[*]

Recently, Kelvin Lancaster (1971) has presented an extension of his earlier approach to the theory of consumer demand (1966a, b). Lancaster's "new approach to consumer theory" is presented as a more general theory of consumer behavior than the traditional one, and is supposed to be "a model very many times richer in heuristic explanatory and predictive power than the conventional model of consumer behavior" (1966a, pp. 154–55). The major contribution of the new approach is dependent on the ability to distinguish between objective and subjective choice and demand theory. However, it is the purpose of this paper to show that this new approach is relevant only under restrictive assumptions in regard to either the nature of the goods consumed or the consumer's utility function.

I

Originally, Lancaster introduced the idea that goods are consumed for the characteristics they possess and they are the objects of consumer preference or utility, with the number of characteristics being greater, equal, or smaller than the number of goods available to the consumer. It is assumed that the characteristics possessed by a good or a combination of goods are, in principle at least, objective and the same for all consumers. The utility derived by the consumer, on the other hand, is subjective and depends on his preference function. Briefly, given the units of measurement any good contains the same amount of any characteristic for all consumers, but each consumer may derive a different level of utility from consuming those characteristics. (See Lancaster (1971, p. 7) and (1966a, p. 134.)

The Lancaster model can be illustrated by considering a simplified choice set of two characteristics A and B, and three goods X, Y, and Z. "Each good gives rise to a vector of the two characteristics, and the consumption technology consists of the activities, consuming each of the goods separately, and consuming them in linear combination" (1966b, p. 16). The imposition of the budgetary constraint on the goods yields the characteristic vectors which confront the consumer. Figure 1 illustrates a case where given the income-price situation, the consumer may obtain points c, d, or e by spending his total income on X, Y, or Z, respectively. The points inside and on the triangle cde represent the bundles of characteristics which are attainable by linear combinations of X, Y, and Z. Lancaster has argued that an efficient consumer will choose combinations on ced, the consumer efficiency frontier. The efficiency frontier changes with relative prices. As the price of Z increases, the characteristics vector which the consumer may obtain

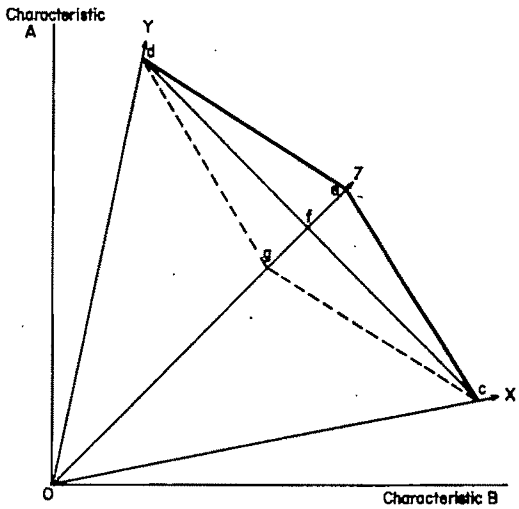


FIGURE 1

---

[*] Senior economist, Gruen, Gruen and Associates; originally written while assistant professor of economics at Temple University. I would like to thank Robert H. Deans for his valuable comments.

by spending all of his income on Z will move toward the origin. If the price of Z increases so that a point left of cd (say, g) becomes the new vector, no combination using Z would be used and cd would become the efficiency frontier, independent of the individual's utility function. Individual preference will determine consumer's equilibrium point on the objective efficiency frontier. The existence of the objective frontier makes it possible to distinguish between two separate effects of any change in income or relative prices: 1) the efficiency substitution effect resulting from a shift in the efficiency frontier, which is objective and independent of the individual consumer's preference function, and 2) the personal substitution effect which depends on the individual's preference function and is similar to the traditional substitution effect. (See Lancaster (1971, pp. 53–65) and (1966a, p. 143).) The existence of the efficiency substitution effect enables the objective observer to make a judgment in respect to the consumer's behavior, even without any knowledge about the latter's utility function.

Such an approach would enable economists to determine, for example, the maximum price above which there should be no demand by efficient consumers.[1] The crucial question is whether the model's assumptions are at least as general as those made by traditional consumer theory, for if the assumptions of the Lancaster model are found to be more restrictive than those of the traditional approach, then it becomes only a special case rather than a general theory of demand. Since the objective efficiency frontier is the foundation of the Lancaster model, it is necessary to examine some of the assumptions required for deriving the efficiency frontier, the meaning and validity of which Lancaster did not analyze.

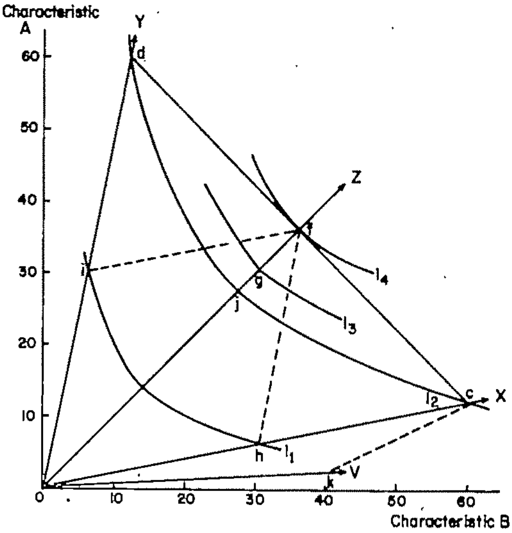


FIGURE 2

II

By considering a choice set of two characteristics A and B and four goods V, X, Y, and Z, and imposing a budget constraint, the characteristic vectors k, c, g, and d in Figure 2 can be derived. According to Lancaster, the efficiency frontier would be cd (i.e., regardless of his preference function the informed consumer should never purchase goods V and Z with the given budget constraint, for by consuming combinations of X and Y the consumer obtains larger quantities of both characteristics than by consuming goods V or Z).[2] Point c, for example, is objectively more efficient than point k. The validity of this argument rests on the assumption that the marginal utility of characteristics may never be negative. Such an assumption of nonnegativity of marginal utility (NNU) is a very restrictive assumption which transcends the assumption that the marginal utility of a good is always positive.[3] While the marginal utility of a good may be high (or even increasing), some of its characteristics may convey disutility. An individual may enjoy eating a sandwich because of its flavor and texture characteristics while experiencing disutility from other characteristics which the sandwich contains, such as its bulk or calorie content. Another individual (or the same person at another time) may feel hungry and therefore enjoy the feeling of consuming bulk but resent the texture. Expectations alone may turn a utility into a disutility. This can be depicted by a person missing his lunch and enjoying to its fullest extent the bulk characteristic contained in a sandwich. Suddenly he recalls that his favorite meal awaits him that evening when he returns from work. Is it still obvious that the utility of the bulk characteristic of the sandwich is continually positive?

---

[1] A consumer facing an income-price situation which gives rise to vectors d, g, and c, for example, is inefficient if he purchases any amount of good Z regardless of his utility function. It can therefore be predicted that efficient consumers will have zero demand for good Z at that price.

[2] Point k in Figure 2 is analogous to point G_1 in Figure 6.3, Lancaster (1971, p. 97), and to point A in Figure 5, Lancaster (1966a, pp. 149–50), when the frontier becomes G_1G_2 and $D'''C$, respectively.

[3] Though the Lancaster model allows for *objectively* negative utility characteristics, it precludes the possibility that a generally desirable characteristic may at some subjective point contain negative marginal utility.

It is difficult to conceive of goods which do not possess characteristics with negative marginal utility while still yielding a net positive marginal utility to the consumer. The common characteristics of bulk, saltiness, sweetness, adventure, usually by their very nature contain positive marginal utility but eventually the marginal utility becomes negative. The point at which it becomes negative is clearly subjective and depends on the consumer's preference function. Consumer *A*, for example, may prefer two spoons of sugar with his coffee while Consumer *B* likes his coffee black. The sugar or sweetness characteristic will be considered a positive marginal utility characteristic (up to two spoons of sugar per cup) by Consumer *A* while containing disutility for Consumer *B*. It is quite possible that Consumer *A* likes his tea without sugar while Consumer *B* likes his tea to be sweet. To accept the *NNU* assumption would very severely restrict the applicability of Lancaster's model. Rejecting this assumption it can be shown that point *c* in Figure 2 is not necessarily preferred by the efficient consumer to point *k*.

Consider the following utility function *u*,

$$u = 2b - 8a \tag{1}$$

where *a* and *b* are the number of units consumed of characteristics *A* and *B*, respectively. Each unit of good *X* contains 1.2 units of *A* and 6 units of *B* and each unit of good *V* contains .2 units of *A* and 4 units of *B*. The total amount of characteristics *A* and *B* consumed with goods *X* and *V* is:

$$a = 1.2x + .2v \tag{2}$$

$$b = 6x + 4v \tag{3}$$

where *x* and *v* are the number of units consumed of *X* and *V*, respectively. By substituting (2) and (3) into (1),

$$U_{x+v} = 2.4x + 6.4v \tag{4}$$

The utility equations show constant utility to scale and constant marginal utility for each good and each characteristic. Figure 2 represents an income-price situation where the consumer's spending of his total income may result in the purchase of either ten units of *X* (point *c*), ten units of *V* (point *k*), or any linear combination of the two (line *kc*). The market price ratio of goods *X* and *V* is one. From (4) we see that $MU_v/MU_x = 6.4/2.4$ and consumer equilibrium requires the purchase of *V* only (point *k*). It is important to note that while characteristic *A* conveys disutility, each good is desirable. It is possible to construct utility functions where the utility of characteristic *A* is negative but *c* is still preferred to *k*. Those functions will be subjective and without such subjective knowledge there is no way to eliminate points such as *k* from the objective efficiency frontier. Relaxation of the *NNU* assumption will not alter Lancaster's model, it will only modify the shape of the efficiency frontier which may now have a positively sloped segment.[4] While the *NNU* assumption is not essential for the Lancaster approach, the derivation of the linear combination frontier is the very basis of his model.

III

Lancaster refers to the analogy of his approach to that of production theory. He assumes that in consumption technology, goods are the inputs and the collection of characteristics is the output. (See Lancaster (1966a, p. 133) and (1966b, p. 14).) This assumption gives the impression that the consumer's objective is to maximize the output of characteristics subject to the budget constraint on the input of goods (and the linear efficiency frontier does just that). The consumer's real objective is to maximize the utility which the characteristics yield, not the amount of characteristics per se. The proper analogy to production theory would be to assume that in consumption technology characteristics are the inputs used to produce the output of utility, while the goods are the processes by which this output is obtained.

---

[4] The frontier in Figure 2 becomes *kcd*. In Figure 6.3, Lancaster (1971, p. 97) it should be $G_3G_1G_2$ and in Figure 5, Lancaster (1966a, p. 149) it becomes $AD'''C$ instead of $D'''C$.

Lancaster's efficiency frontier is derived by a linear combination which is similar to the linear combination isoquant derived by Robert Dorfman, p. 805. The linear combination isoquant indicates the aggregate level of output obtainable by utilizing several production processes. Any given point on the isoquant indicates a given amount of each input used by the various processes. If the same total quantities of the inputs were to be used by a different process the resulting level of output would be incomparable to that of the linear combination isoquant.[5] Lancaster's approach, however, is based on just such comparisons. The derivation of the efficiency frontier presupposes the ability to compare the utility (output) of a linear combination of two separate goods (processes) with that of a third good (process). According to Lancaster, point *f*, which is obtained by a linear combination of vector *h* of good *X* and vector *i* of good *Y*, is preferable to point *g* which might be obtained by consuming good *Z* alone. Although point *f* represents a larger amount of total characteristics, the conclusion that it is preferable to point *g* (which is obtained by a different process) requires either one of the two following assumptions: 1) the utility independence of characteristics per consumption unit (*UIC*), or 2) the "mixability" of goods per consumption unit (*CU*). The *UIC* assumption means that while total utility of one characteristic may depend on the total amount of the other characteristics, it is independent of the ratio of characteristics per unit of consumption. The mixability assumption means that two different goods (such as *X* and *Y*), each containing a different ratio of characteristics, can be mixed to yield the same ratio of characteristics as a third good (*Z*) per unit of consumption.

The *CU* concept is of great importance to consumer theory. The consumption of goods and services is very seldom instantaneous. The utility derived from the consumption of a good depends on the length of time over which it is consumed and the manner in which it is consumed. The utility derived from drinking an average amount of one fifth of whiskey per week, for example, will depend on the time distribution of its consumption. Drinking a full fifth every Sunday and none the rest of the week is most likely to yield a different level of utility per fifth a week than having a couple of drinks each day. The utility derived from consuming an apple is a function of how the apple is to be consumed: consuming the apple in five minutes by taking ten equal-size bites every thirty seconds is likely to yield a different level of utility than taking the same size bite every thirty minutes, or consuming it by taking equal bites every five seconds for five minutes. The theory of consumer behavior implicitly assumes that the consumer distributes his consumption of every good in such a "time-manner" combination that will maximize his utility, i.e., the consumer is assumed to be time-manner efficient. The utility of any individual is, therefore, a function not only of the quantities of each good he consumes but also of the time-manner coefficient of each good. The time-manner coefficients are in turn a function of the quantities consumed. A consumer having only one apple (or drink) a day will probably have a different optimal time-manner coefficient than the one he would have if the price were to decline so that he would be consuming three apples (or drinks) a day. The *CU* refers to the unit of consumption which maximizes consumer's utility. In the above example it would be measured in terms of the optimal bite-size of the apple. The *CU* is clearly subjective. It is a function of taste, the quantity consumed of each good, expectations, etc.

---

[5] The output of a process utilizing ten units of labor and ten units of capital, for example, is unlikely to be equal to the sum of the outputs produced by one process utilizing ten units of labor and zero capital and another process utilizing ten units of capital and zero units of labor.

The mixability assumption is valid only in respect to a very small portion of goods consumed. The validity of the model as a general theory must rest on the *UIC* assumption. This implies that the individual is indifferent to whether he consumes one very sweet apple and one without any sweetness (but otherwise identical) or two mildly sweet apples, as long as the sum of sweetness and all other characteristics are the same. Or one rare and cold steak with two units of spices and one very hot well-cooked steak without any spices should give the same satisfaction as two warm medium-cooked steaks each containing one unit of spices. Or driving one hour in a car which is uncomfortable (or unsafe) and one hour in a car which is very comfortable (or safe) gives the same utility as driving two hours in a moderately comfortable (or safe) car, so long as the total amount of each characteristic is the same.[6] Or only inefficient consumers (or people who enjoy making them) would consume salads. The total amount of characteristics is the same whether the fruits or vegetables are consumed separately or as a salad. The preparation of a salad requires time and effort which is equivalent to paying a higher price for the salad than for the fruits or vegetables of which it is composed. Therefore, the vector representing the salad should be below Lancaster's efficiency frontier.

The above examples clearly indicate the invalidity of the *UIC* assumption. Consumers do care about the distribution of each characteristic per consumption unit (*CU*). The *CU* differs from good to good and depends on the subjective taste of the consumer. The utility of most goods and most characteristics depends on the ratio of characteristics per *CU* and not merely on their total amount. The quality of combining different characteristics in different ratios per *CU* is what gives goods their uniqueness. Upon relaxing the *UIC* assumption as a general phenomenon, the Lancaster analysis breaks down. Consider a total utility function which is a linear combination of the utility derived by the different processes (goods), and the utility function of each process is of the Cobb-Douglas type.

$$U = \sum_{i=1}^{n} U_i \tag{5}$$

$$U_i = 10a_i^{.5}b_i^{.5} \tag{6}$$

where U_i is the utility of the ith process (good) and a_i and b_i are the number of characteristics A and B, respectively, in the ith process. The utility function in our example[7] is:

$$U = U_X + U_Y + U_Z \tag{7}$$

assuming

$$a = 1.2x + 6y + 3z \tag{8}$$

$$b = 6x + 1.2y + 3z \tag{9}$$

where x, y, and z are the number of units consumed of goods X, Y, and Z, respectively. Point f in Figure 2 is obtained by consuming five units of X and five units of Y. The total number of characteristics is 36 units of A and 36 units of B. Point g is obtained by consuming ten units of Z. The total amount of A is 30 and the total amount of B is 30. From the utility functions and equations (8) and (9) it is possible to compare the utilities at points f and g.

$$\begin{aligned} u_f = u_x + u_y &= 10(1.2x)^{.5}(6x)^{.5} \\ &\quad + 10(6y)^{.5}(1.2y)^{.5} = 268.3 \end{aligned} \tag{10}$$

$$u_g = u_z = 10(3z)^{.5}(3z)^{.5} = 300 \tag{11}$$

---

[6] The last example differs from the one given by Lancaster (1966b, p. 21) for products that cannot be utilized in combination. Here it is assumed that the consumer may use them in combination by renting the cars.

[7] To simplify we assume that the utility of each characteristic is positive, which excludes good V from any consideration.

Point g which is obtained by consuming Z is preferable in this case to point f which is achieved through a linear combination of X and Y. The comparison between the utility at f and that of g is therefore subjective since it assumes a particular type of utility function—the *UIC* assumption.

The indifference curves in Figure 2 were derived from equation (6) where only one single process is utilized.

$$U = 10A^{.5}B^{.5} \tag{12}$$

and along any indifference curve

$$dU = 0 = \frac{5B^{.5}}{A^{.5}}dA + \frac{5A^{.5}}{B^{.5}}dB \tag{13}$$

and the slope of each indifference curve is

$$\frac{dA}{dB} = -\frac{A}{B} \tag{14}$$

The indifference curves in Figure 2 are, therefore, convex to the origin.

Each indifference curve is meaningful only when each point on the curve is achieved by a single process. Any point on the same indifference curve may reflect a different level of utility if obtained through a linear combination rather than a single process. Point f, for example, is on indifference curve I_4 only if obtained by consuming good Z alone. The utility at point f if obtained through a linear combination (of points h and i) is, in the above example, the same as that of points c or d, i.e., the same as that of indifference curve I_2 which is derived for a single process. No objective judgment is possible about the utility ranking of various bundles of characteristics when they are achieved through different linear combinations. The Lancaster linear combination line dc is a "characteristics possibility frontier" rather than an efficiency frontier. Nothing can be said in terms of consumer efficiency unless the goods are known to be mixable or the consumer's utility function is given. The Lancaster efficiency frontier is objective only under certain restrictive assumptions. The "new approach" thus becomes an interesting and important special case of consumer choice rather than a general model of consumer demand.

REFERENCES

R. Dorfman, "Mathematic or Linear Programming: A Nonmathematical Exposition," *Amer. Econ. Rev.*, Dec. 1953, *43*, 795–828.

K. Lancaster, (1966a) "A New Approach to Consumer Theory," *J. Polit. Econ.*, Apr. 1966, *84*, 132–57.

———, (1966b) "Change and Innovation in the Technology of Consumption," *Amer. Econ. Rev. Proc.*, May 1966, *56*, 14–23.

———, *Consumer Demand: A New Approach*, New York 1971.

A Note on Uncertainty and the Evaluation of Public Investment Decisions

By ESTELLE JAMES*

The optimal handling of risk by the government continues to be a subject of controversy. Some economists maintain that the government should base its decisions on prices prevailing in the market, while others hold that the appropriate social risk premium is lower than the private. Early proponents of the latter viewpoint pointed to the advantages of diversification and risk pooling from which large organizations such as the government benefit. In a recent major contribution, Kenneth Arrow and Robert Lind have added another justification: by spreading the risk across large numbers of people, the government reduces the overall cost of bearing risk to zero in the limit.[1]

Risk pooling applies in the many-project case, regardless of the number of investors. Risk spreading applies in the many-investor case, regardless of the number of projects. The purpose of this paper is to explore the relationship between these two arguments and their implications for the much debated issue of whether investment projects undertaken by the government should be evaluated individually or as a group.

Section I analyzes the risk premium as arising from two characteristics of a gamble: its scale and its dispersion. The Arrow-Lind argument for a low social risk premium stems from a scale effect, while the risk-pooling argument stems from a dispersion effect. For both, there may be an inconsistency between the piecemeal and global evaluation of projects, but the nature of the conflict is reversed in these two cases. The effect on dispersion of risk pooling may lead the government to *accept a group* of projects which were *individually rejected.* On the other hand, the scale effect may lead the government to *reject an entire group* of projects which were *accepted individually* due to risk spreading. This inconsistency is avoided and piecemeal and global decisions will coincide when the scale effect (risk spreading) dominates in the evaluation of individual projects and the dispersion effect (risk pooling) dominates in the group evaluation. Section II develops this theme further, using diagrammatic techniques.

Section III investigates the necessary and sufficient assumptions for another Arrow-Lind finding—that if the time-state preference approach is used, the value of an investment to the individual depends only on its expected return. This finding is shown to hold only under rather strong restrictive conditions which were not explicitly specified in their paper.

I

A. *Risk Spreading and the Scale Effect for Individual Projects*

The Arrow-Lind argument may be summarized as follows: Assume there are n identical people in an economy, each with disposable income represented by A. A project is under consideration which would yield a return of $\overline{B}+X$, where $\overline{B}$ is the expected net benefit and X the random part with mean $u_X=0$. The risk cost attached to this project if undertaken by one person (i.e., the amount he would willingly pay to avoid the risk represented by X) may be denoted k_X, and for $k_X>\overline{B}$, he would not accept the project. If the risk is shared among n people, each person's part of the return becomes $(\overline{B}+X)/n$. The corresponding risk cost (i.e., the amount an individual would pay to avoid the risk represented by X/n) becomes $k_{X/n}$ for one person and $nk_{X/n}$ for all, while the expected social return remains $\overline{B}$. For large n, $nk_{X/n} < k_X$, and further, $nk_{X/n} \rightarrow 0$ as $n \rightarrow \infty$. Therefore, we may have $k_X > \overline{B} > nk_{X/n}$, in which case the project should be undertaken collectively, although it would not be undertaken privately.

---

* Professor of economics, State University of New York, Stony Brook. I would like to thank my colleagues, R. Kihlstrom, E. Neuberger, and M. Nienhaus for their helpful comments on an earlier version of this paper.

[1] See Arrow and Lind for the risk-spreading justification. The private market argument is stressed by Jack Hirshleifer and the risk pooling argument was set forth by Paul Samuelson and William Vickrey, among others.

We may restate this conclusion in terms of a scale effect and a dispersion effect on risk costs, two concepts which will be developed further in this note. Define the scale of a gamble in terms of the maximum possible positive and negative values of the random variable, i.e., its range. Two random variables Y and Z with means $u_Y = u_Z = 0$, maximum positive values Y_1 and Z_1 and minimum values Y_2 and Z_2 are said to have the same scale if $Y_1 - Y_2 = Z_1 - Z_2$. Define the relative dispersion of a gamble in terms of the probabilities π attached to any two values of the random variable. Two random variables Y and Z are said to have the same relative dispersion if $\pi_{qY_1} = \pi_{qZ_1}$ and $\pi_{qY_2} = \pi_{qZ_2}$, where q can take on any value between 0 and 1 and π_{qY_1} signifies the probability that $Y = qY_1$.

In the Arrow-Lind example, where the random variable simply becomes X/n instead of X, the scale of the gamble has been multiplied by $1/n$ while the relative dispersion has remained the same, i.e., $\pi_{qX_1} = \pi_{q(X/n)_1}$ and $\pi_{qX_2} = \pi_{q(X/n)_2}$. In such a case, Arrow-Lind have shown the risk premium an individual imputes to the smaller gamble is less than $1/n$th his risk premium for the larger gamble. The expected return, however, has only decreased from $\overline{B}$ to $\overline{B}/n$; i.e., net benefits fall in proportion with scale while risk costs fall faster than scale. This *pure scale effect* on risk costs accounts for the advantages of risk spreading demonstrated by Arrow-Lind.

Arrow-Lind prove their point mathematically, but it is worthwhile inquiring into its intuitive rationale. Risk aversion is usually assumed to stem from diminishing marginal utility of income, so that in a fair gamble the utility of the potential income gain is less than the utility of the potential loss. For example, in Figure 1 the slope of the utility function over AY_1 is less than that over Y_2A; let us call w_{kY} the utility difference associated with the fair gamble of scale $Y_1 - Y_2$ leading to risk premium k_Y. If we now double the scale of the gamble, say from $Y_1 - Y_2$ to $Z_1 - Z_2$ (with its relative dispersion unchanged), we find an even greater difference between the utilities of the incremental income gain and loss, i.e., the difference in slope between Y_1Z_1 and Z_2Y_2 is larger than the difference between AY_1 and Y_2A. Hence, the total utility difference has more than doubled; $w_{kZ} > 2w_{kY}$. In general, when the scale of a gamble is multiplied by a factor p, the utility difference changes more than p. This in turn suggests that the risk premium changes more than p.[2] Applying this analysis to the Arrow-Lind case, when scale is multiplied by $1/n$, $X \rightarrow X/n$, $\overline{B} \rightarrow \overline{B}/n$ but $k_X \rightarrow k_{X/n} < k_X/n$, so for large n, $k_{X/n} < \overline{B}/n$ and $nk_{X/n} < \overline{B}$ even if $k_X > \overline{B}$.

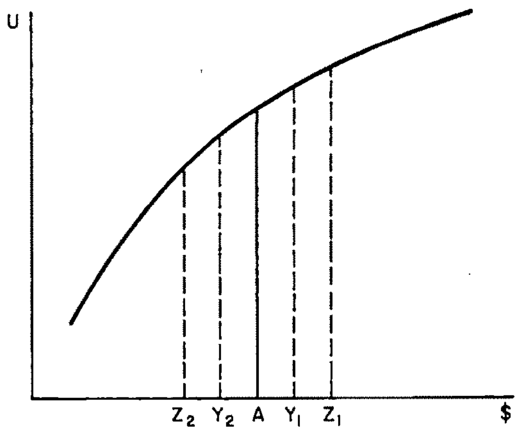


FIGURE 1. UTILITY, RISK AVERSION, AND SCALE

It should be noted that the same effect could be achieved by a single individual without any collective action for projects characterized by perfect divisibility and constant returns. Thus, risk spreading leads to a divergence between social and private decisions about risk projects only in the absence of these characteristics.

---

[2] This explanation for the scale effect also holds in the case of concave utility functions and preference for gambling: the premium such an individual would be willing to pay for the opportunity to gamble would then increase faster than the scale of the gamble.

B. *Risk Spreading, Scale, and Global Evaluations*

While the advantages of risk spreading hold when each (indivisible) project is evaluated individually, it seems reasonable to ask whether the result arrived at by marginal decision making is globally optimal. To explore this question, let us assume that m projects are under consideration, each having expected benefit $\overline{B}$ and random return X with mean $u_X=0$. The risk cost to a single person undertaking one of these projects is k_X; if shared among n identical people, the risk cost per person is $k_{X/n}$. As above, $k_X>\overline{B}>nk_{X/n}>0$ for large but finite n; thus, each project is acceptable on a shared basis only, and all m projects are undertaken collectively.[3] Is society satisfied with this result overall?

Suppose first that the returns of these projects are perfectly correlated; that is, they depend on the state of the world such that for any given state all projects will yield exactly the same return. The total expected return for each person from the m projects taken as a group is now $m(\overline{B}/n)$, the random part is $m(X/n)$, and the corresponding risk premium is $k_{mX/n}$. If $m=n$, these expressions may be simplified: $m(\overline{B}/n)=\overline{B}$; $m(X/n)=X$; and $k_{mX/n}=k_X$. That is, the risk attached by the individual to the bundle of n shared correlated projects is k_X, exactly the same as the risk associated with one whole project. We know that the latter is greater than $nk_{X/n}$, which we may now interpret as the sum of the separate risk premiums imputed by the individual to each of the n shared projects. We also know that k_X exceeds the expected return $\overline{B}$. The larger number of piecemeal collective investment decisions has cancelled the scale effect. Consequently, when the individual reevaluates his position globally, he finds himself in exactly the same choice situation he was in initially when, quite rationally, he preferred not to invest.

There exists, in other words, a possible inconsistency between marginal and global decisions, due to the scale effect. Each shared project is acceptable on its own, but if the entire bundle is reconsidered, a typical individual would prefer to eliminate them all. The larger m is relative to n, the larger $k_{mX/n}$ is relative to $mk_{X/n}$ and the greater the likelihood that marginal decisions based on risk spreading will not lead to a preferred outcome globally.

C. *Risk Pooling and the Relative Dispersion Effect*

Is there any way out of this dilemma? In some cases yes, but only if the projects are independent so that there are gains from risk pooling, which changes the relative dispersion of returns.

Suppose then that the returns to the m projects are independent, rather than being perfectly correlated. If $m=n$, the expected return to each person remains $n(\overline{B}/n)=\overline{B}$; the random term we shall denote as $\sum(X/n)$. We cannot simply say that $\sum(X/n)=X$, since the relative frequency distributions of these variables differ. Specifically,

$$\pi_{q(\Sigma X/n)_1}>\pi_q X_1 \quad \text{and} \quad \pi_{q(\Sigma X/n)_2}>\pi_q X_2$$

for q near 0 and vice versa for q near 1. That is, the probability attached to $\sum(X/n)$ is higher for values near $u_X=0$ and lower for values far from zero than is the probability attached to X. The scale of $\sum X/n$ (n pooled shared projects) is, however, the same as that of X (a single nonshared project), since $(\sum X/n)_1=X_1$ and $(\sum X/n)_2=X_2$. Thus we have isolated the *pure dispersion effect*, ordinarily associated with risk pooling. As discussed in the risk-pooling literature, the denser frequency distribution of $\sum X/n$ means that $k_{\Sigma X/n}<k_X$.

---

[3] We are assuming that the risk cost $k^i_{X/n}$ is the same for all projects $i=1, \ldots m$ and, further, does not change as expected income changes when some of the projects have been undertaken. This simplification is made for expositional purposes; the argument still holds when $k^i_{X/n}$ differs or changes, so long as $k^i_X>\overline{B}$ but $nk^i_{X/n}<\overline{B}$ for all i. Likewise, the assumptions that $\overline{B}$ is the same for all i and later that $m=n$, are made for expositional convenience and are not essential to the main result—that there may be an inconsistency between global and marginal decision making due to the scale effect. The assumption of perfect correlation is, however, crucial as will be shown later.

D. *Pooling Shared Risks: Scale and Dispersion Effects Combined*

We have found that because of the pure scale effect, $k_X > \bar{B} > nk_{X/n}$. We have also found that because of the pure dispersion effect, $k_{\Sigma X/n} < k_X$. It remains now to compare $k_{\Sigma X/n}$ with $nk_{X/n}$ and $\bar{B}$.

In making this comparison, both the scale and dispersion effects must be taken into account. As we have observed, the scale of $\sum X/n$ (n pooled shared projects) is n times the scale of X/n (a single shared project) and therefore $k_{\Sigma X/n}$ may be greater than $nk_{X/n}$ and $\bar{B}$. In this sense, risk pooling introduces a scale effect, which increases the risk premium. On the other hand, $\sum X/n$ has a denser relative probability distribution than X/n and on this account $k_{\Sigma X/n}$ may be smaller than $\bar{B}$ and $nk_{X/n}$. While both $nk_{X/n}$ and $k_{\Sigma X/n} \to 0$ as $n \to \infty$, the former because of the scale effect and the latter because of the dispersion effect, we may for a given finite n have $k_{\Sigma X/n}$ greater or smaller than $nk_{X/n}$ and $\bar{B}$, depending on the relative strength of these two effects.

If the scale effect dominates, we will have $k_X > k_{\Sigma X/n} > \bar{B} > nk_{X/n}$. Global and marginal decisions still conflict and an entire group of individually acceptable projects may be rejected globally. If, on the other hand, the dispersion effect is sufficiently strong, we will have $k_X > \bar{B} > k_{\Sigma X/n} \lessgtr nk_{X/n}$. Since $k_{\Sigma X/n}$ is a declining function of n when the projects are independent, we can always find an n large enough to bring about the latter case and avoid the inconsistency between global and marginal decisions described above.[4]

II

This analysis may be illustrated diagrammatically, as follows. For any given person, the risk cost k attached to a gamble represented by a random variable Y with mean $u_Y = 0$ depends on two characteristics of Y: its scale and its relative dispersion. If the latter is held constant, k_Y is an increasing function of scale and the second derivative is also positive. If the distribution of Y changes, we get an entirely different function which for "better" distributions lies completely below the first; in Figure 2, $k^B < k^A$ for any given scale. Thus, we have a whole family of curves depicting k as a function of scale, each curve associated with a different relative dispersion of Y.

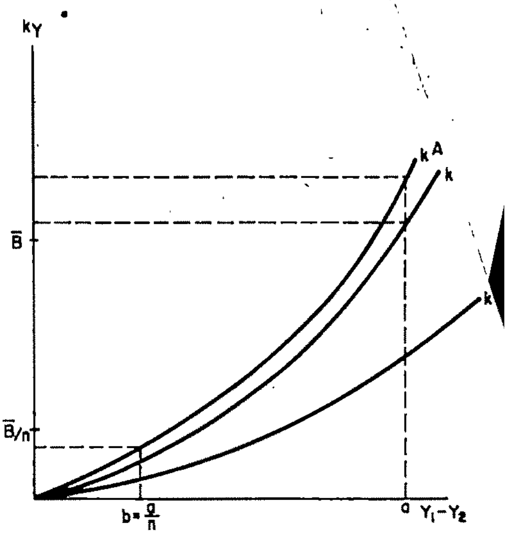


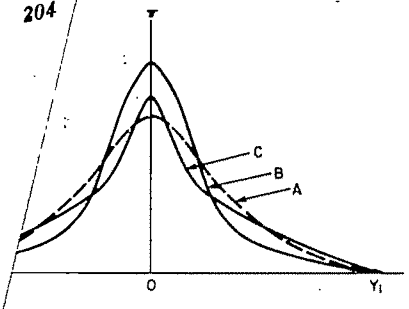
FIGURE 2. RISK COST AS A FUNCTION OF SCALE AND RELATIVE DISPERSION

While scale can easily be quantified as $Y_1 - Y_2$, the ordering of the relative dispersion depends on subjective rather than objective evaluations. That is, different individuals may have different preference orderings for the various possible distributions and even if they agree on ranking they may disagree on the absolute value of the change. For example, in Figure 3 some individuals might rank distribution A over distribution C and others vice versa. A shift from A to B, on the other hand, would be deemed an improvement by all risk averters, but they might disagree on the size of the gain.

Now, in the initial situation described by Arrow-Lind, the investor faces random variable X with scale a and relative dispersion A such that $k_X > \bar{B}$; the project is not accepted (see Figure 2). Risk spreading among n people moves him to random variable X/n with scale $b = a/n$ and dispersion

---

[4] For a closely related discussion of $k_{\Sigma X/n}$ see Richard Kihlstrom and Mark Pauly.



3. DIFFERENT RELATIVE DISPERSIONS FOR SAME SCALE

which $k_{X/n}<\overline{B}/n$; the project is ac- When n perfectly correlated pro- are undertaken and shared, the scale ts to a, the dispersion still unchanged; $_{/n}=k_X>\overline{B}$ and the investor regrets his ecision to accept the entire bundle.

If instead the projects are statistically independent, the pooling process changes the distribution and hence places the investor on another curve, k^B or $k^{B'}$ with random variable $\sum X/n$ at scale a. Since the new relative dispersion is unambiguously better than the old, both k^B and $k^{B'}$ are lower than k^A. However, in the latter case the scale effect dominates at a so $k_X>k_{\Sigma X/n}>\overline{B}>nk_{X/n}$. If on the other hand the individual has a stronger preference for B over A, k^B becomes relevant; the dispersion effect outweighs the scale effect sufficiently to make $k_X>\overline{B}>k_{\Sigma X/n}\lesseqgtr n k_{X/n}$. The larger the number of pooled independent projects, the more likely that this will be the case. In other words, piecemeal decisions based on risk spreading will lead to a correct global outcome if and only if sufficiently large gains from risk pooling are also present, when the investments are reevaluated as a group.

III

The Arrow-Lind paper also proved that if the time-state preference approach is used and a perfect market exists for contingent claims in each possible state of the world, the value of an investment to an individual depends only on its expected return. This runs counter to the usual presumption that the value of a project is a function of the state-distribution of its returns and the utility functions, subjective probabilities and prices associated with different states of the world.

Arrow-Lind derive their surprising conclusion by moving from a situation where each state is treated separately to a situation where states are aggregated according to utility functions and returns. A crucial step in this transition involves their statement that "... the allocation X_{itq} depends only on t, i.e., is the same for all states in E_t for any given t" (p. 369) so that $\partial U_{is}/\partial X_{isq}$ in equation (3) may be replaced by $\partial U_{it}/\partial X_{itq}$ in equation (9).[5] Upon closer examination, it appears that this procedure is valid if and only if two additional assumptions are made about initial endowments and subjective probabilities.

The following observations are suggestive of this finding. If $X_{irq}=X_{isq}$ for two states r and s in a given E_t and for all individuals $i=1,\dots I$, then $\sum_i X_{irq}=\sum_i X_{isq}$. Since by definition, the total consumption claims in each state cannot be changed by trading, $\sum_i X_{irq}=\sum_i \overline{X}_{irq}=\sum_i \overline{X}_{isq}=\sum_i X_{isq}$. *That is, the total initial endowments* $\sum_i \overline{X}_{irq}$ *and* $\sum_i \overline{X}_{isq}$ *would have to be the same for all states with the same utility function; the total number of available consumption units could not differ for two states which happen to have the same* U_{is}.

---

[5] In order to save the reader's time, the symbols used in Arrow-Lind paper and in this section of my note are defined below:

X_{isq} = the equilibrium claim of individual i in state s in time period q+1.
$\overline{X}_{isq}$ = the initial (consumption claim) endowment of individual i in state s in time period q+1.
E_t = a partition of all states of the world such that, for all s in any given E_t all utility functions U_{is} are the same for any individual i; this partition is assumed to be the same for all individuals.
F_u = a partition of all states of the world such that, for all s in any given F_u, the returns on a particular investment project are the same.
E_{tu} = the set of all states of the world which lie in both E_t and F_u.
P_{sq} = the present value of a single commodity unit at time q+1, given state s.
π_u = the probability of F_u, assumed to be the same for all individuals.
π_{is} = the subjective probability that individual i assigns to the occurrence of state s.

Also, if $X_{irq}=X_{isq}$ and $U_{is}=U_{ir}$, then $P_{sq}/P_{rq}=\pi_{is}/\pi_{ir}$, $i=1,\dots I$, by Arrow-Lind's equation (4) which states that

$$\frac{P_{sq}}{P_{rm}}=\frac{\pi_{is}}{\pi_{ir}}\,\frac{\partial U_{is}/\partial X_{isq}}{\partial U_{ir}/\partial X_{irm}}$$

That is, all people must have equal, exogenously given, subjective relative probabilities for any two states in a given E_t. Neither of these assumptions is explicitly specified in the Arrow-Lind model.

We can show more formally that the equal endowments and relative probabilities assumptions constitute necessary conditions for the desired results as follows: If subjective relative probabilities differ among people, relative marginal utilities must also differ in equilibrium, by equation (4). Hence $X_{isq}\neq X_{irq}$, $\partial U_{is}/\partial X_{isq}\neq\partial U_{ir}/\partial X_{irq}$ and both cannot be replaced with $\partial U_{it}/\partial X_{itq}$ for at least some individuals. If relative probabilities are the same for all people but $\sum\overline{X}_{isq}>\sum\overline{X}_{irq}$, $(\partial U_{is}/\partial X_{isq})/(\partial U_{ir}/\partial X_{irq})$ must be the same for all by equation (4), so all people must end up with $X_{isq}>X_{irq}$ and $\partial U_{is}/\partial X_{isq}<\partial U_{ir}/\partial X_{irq}$; therefore we could not replace both with the expression $\partial U_{it}/\partial X_{itq}$.

The equal endowments and relative probabilities assumptions are also sufficient conditions for these results. Suppose that for s and r in a given E_t, $\sum\overline{X}_{irq}=\sum\overline{X}_{isq}$ and π_{is}/π_{ir} is the same for all people. Because of the latter, $(\partial U_{is}/\partial X_{isq})/(\partial U_{ir}/\partial X_{irq})$ must also be the same for all people since they face the same prices. If $\partial U_{is}/\partial X_{isq}\lesseqgtr\partial U_{ir}/\partial X_{irq}$, then $X_{irq}\lesseqgtr X_{isq}$ for all i so that $\sum X_{irq}=\sum\overline{X}_{irq}\lesseqgtr\sum X_{isq}=\sum\overline{X}_{isq}$. This, however, would contradict the equal endowments assumption. Therefore, to satisfy this assumption, $X_{isq}=X_{irq}$ for all i, $\partial U_{is}/\partial X_{isq}=\partial U_{ir}/\partial X_{irq}$, and $P_{sq}/P_{rq}=\pi_{is}/\pi_{ir}$.

Thus, if these restrictions are made on the magnitudes of the initial endowments and the subjective probabilities, the equilibrium X_{isq} must be identical for all states s within a given E_t, $\partial U_{is}/\partial X_{isq}$ may be replaced by $\partial U_{it}/\partial X_{itq}$ and the market-clearing set of relative prices must equal the relative probabilities. If these restrictions are not made, the allocation of X_{isq} must vary for different states in a given E_t, $\partial U_{is}/\partial X_{isq}$ could not simply be replaced by $\partial U_{it}/\partial X_{itq}$, and relative prices depend on both probabilities and marginal utilities. While Arrow-Lind specify various assumptions regarding the uniformity of π_{it} for all people and all t and the independence of returns on this investment from previous returns, the equal endowments and relative probability requirements are additional. They constitute both necessary and sufficient conditions for the Arrow-Lind results.

The time-state preference approach suggests that the different state preferences and their prices must be used in evaluating an investment since the returns may be concentrated in states with 1) low (high) initial endowments, or 2) high (low) utility functions, or 3) high (low) probabilities and therefore high (low) time-state prices. The Arrow-Lind analysis has shown that when the first two possibilities are assumed away and certain restrictions are placed on the probability structure only the expected return need be taken into account in making a risky investment decision.

REFERENCES

K. J. Arrow and R. C. Lind, "Uncertainty and the Evaluation of Public Investment Decisions," *Amer. Econ. Rev.*, June 1970, *60*, 364–78.

J. Hirshleifer, "Investment Decision Under Uncertainty: Applications of the State Preference Approach," *Quart. J. Econ.*, May 1966, *80*, 252–77.

R. Kihlstrom and M. Pauly, "The Role of Insurance in the Allocation of Risk," *Amer. Econ. Rev. Proc.*, May 1971, *61*, 371–79.

P. A. Samuelson and W. Vickrey, "Discussion," *Amer. Econ. Rev. Proc.*, May 1964, *54*, 88–96.

Tax Policy and Investment: An Analysis of Survey Responses

By ROBERT EISNER AND PATRICK J. LAWLER*

Economic policy in the United States in recent years has included tax measures designed to affect the level of business investment. These have taken several forms: accelerating rates of tax depreciation on capital goods, thus lowering the present value of expected tax liabilities and actually decreasing annual tax payments; tax "credits" amounting to subsidies for the purchase of equipment; and alterations in business income tax rates. With additional acceleration of tax depreciation in the "Asset Depreciation Range" system and reenactment of an equipment tax credit in 1971, and recent proposals for suspension and then for increases in the credit, the issues are particularly current.

A number of analyses have attempted to estimate the effects of investment tax incentives by incorporating their presumed implications in more general variables, such as the cost or "rental price" of capital, and estimating the parameters of these more general variables.[1] In some instances attempts have been made to estimate effects more directly, either entering tax rates separately or isolating the specific changes in more general variables which have been due to the tax measures.[2] Our efforts here are directed primarily at what business respondents in McGraw-Hill surveys *said* a number of tax measures would do or had done to their anticipated or actual capital expenditures and comparing these with several econometric projections. We shall also report briefly on inconclusive results of inclusion in general investment functions of the survey responses as to anticipated or actual effects of the tax measures on expenditures.

I. The Survey Data

The current analysis utilizes McGraw-Hill capital expenditure surveys of firms in nonfinancial industries. The following questions in the 1963, 1964, 1965, 1966 (fall), and 1968 surveys deal with the effects of tax measures on investment.[3]

In the survey for the spring of 1963, McGraw-Hill included the question, "Of the total amount you now plan to invest in 1963, roughly how much is due to: a) New depreciation schedules? ___ %; b) Tax credits for new equipment? ___ %." In the survey of the spring of 1964, a similar question referred to plans to invest in 1964.

The spring survey of 1965 asked, "Of the amount you invested in 1964, roughly how much more did your company spend than it would have because of: a) Reduction in corporate tax rate? ___ %; b) Revised depreciation schedules? ___ %; c) Tax credits for new equipment? ___ %."

Then in the fall of 1966, McGraw-Hill asked, "How much have you reduced your 1967 capital investment plans because of suspension of: a) Investment tax credit? \$___; b) Accelerated depreciation on structures? \$___."[4]

---

* Professor of economics, Northwestern University, and instructor, University of North Carolina at Asheville, respectively. Eisner is a member of the research staff of the National Bureau of Economic Research. Underlying research has benefitted from financial support of the National Science Foundation. None of these institutions, of course, is in any way responsible for the contents or views expressed in this paper. We are most grateful to Robert Coen for comments on an earlier draft.

[1] See, for example, Charles Bischoff, Robert Coen, Robert Hall and Dale Jorgenson, and Lawrence Klein and Paul Taubman in the collection of papers edited by Gary Fromm; also Coen and Jorgenson.

[2] See Eisner (1969), in particular.

[3] The authors are indebted to the Department of Economics of the McGraw-Hill Publishing Company for making available the actual questionnaires used and individual firm responses, with firms coded by number to preserve the confidentiality of the data.

[4] We divided these responses in dollars by anticipated

TABLE 1—SURVEY RESPONSES AND INDEPENDENT ECONOMETRIC ESTIMATES[a] OF THE EFFECTS OF TAX POLICY ON CAPITAL EXPENDITURES

(Shown in Percent)

		Percent Effects on Capital Expenditures					
		Mean of Survey Responses[b]		Econometric Estimates			
	Year	Simple	Weighted	Hall and Jorgenson	Bischoff	Klein and Taubman	Coen
Proportion of anticipated capital expenditures due to:							
New depreciation guidelines	1963	4.2 (339)	1.9 (277)	2.1 *(ex post)*	0.2[d]	0.3	1.9
Equipment tax credit	1963	2.8 (337)	1.4 (274)	4.4	0.6[d]		
New depreciation guidelines	1964	2.8 (301)	1.3 (273)	—	—	—	—
Equipment tax credit	1964	1.1 (299)	0.8 (271)	—	—	—	—
Proportion of actual capital expenditures due to:							
Corporate income tax reductions	1964	0.8 (275)	0.5 (245)	−0.4[c]	0.05	—	3.5
New depreciation guidelines	1964	1.5 (273)	0.9 (243)	2.2	1.0[d]	0.9	
Equipment tax credit	1964	1.3 (272)	1.1 (242)	6.8	3.9[d]		
Proportion anticipated capital expenditures reduced due to:							
Suspension of accelerated depreciation on structures	1967	−0.7[e] (331)	−0.1[e] (332)	−1.4	—	−2.8	—
Suspension of equipment tax credit	1967	−2.2[e] (331)	−0.7[e] (339)	−6.8	—		
10 percent corporate tax surcharge	1968	−1.6 (281)	−0.2 (160)	—	—	—	—

[a] For data base, see Appendix.

[b] Number of responses in parentheses.

[c] Assumed no effects on output.

[d] Bischoff's estimates relate only to equipment. If the depreciation guidelines and equipment tax credit had no effect on expenditures for structures the implicit estimate of their percent effect on all capital expenditures would be roughly two-thirds of the figure shown.

[e] McGraw-Hill reported, as to the aggregate, that suspension of accelerated depreciation on structures and of the equipment tax credit had reduced fall 1966 capital expenditure plans by $1.5 billion out of a total of some $63.4 billion (Press release, February 10, 1967, "McGraw-Hill's Special Check-up on 1967 Investment Plans") or about 2.4 percent, a somewhat higher figure than the weighted mean total of 0.8 percent but less than the simple mean total of 2.9 percent calculated from our partial sample.

Finally in the spring of 1968, we have the question, "If the proposed 10% surcharge on corporate income taxes is enacted, how much do you estimate this will reduce your capital spending plans for 1968? ___ %."

Table 1 presents information regarding all of the responses available. Working with the raw means of survey answers as to the percent effect of the various measures on anticipated or actual expenditures, we find respondents generally indicating small movement in the presumably correct directions, and usually somewhat larger effects from

---

capital expenditures reported in answer to an accompanying question, thus securing observations in ratio form comparable to those relating to questions on the other surveys.

liberalized depreciation than from the tax credits for new equipment.[5] Thus the ununweighted mean proportion of anticipated investment in 1963 attributed to the new depreciation schedule was 4.2 percent while that attributed to tax credits for new equipment was 2.8 percent. The similar question in 1964 again shows a larger amount of anticipated investment attributed to liberalized depreciation than to the tax credit but, interestingly, considerably smaller proportions for both than in the previous year.

That this is so, despite the cumulative gains from continued accelerated depreciation and the fact that 1964 investment should have reflected the lagged results of earlier as well as current stimuli, raises some question as to the internal consistency of the responses. This concern may be heightened by the *ex post* response in early 1965 regarding actual 1964 expenditures. For the proportion then attributed to the revised depreciation schedule was only 1.5 percent, little more than half of the 2.8 percent which had been indicated in previous anticipations of those expenditures and little more than a third of the 4.2 percent mean figure noted with regard to the 1963 anticipations.

The suspension of accelerated depreciation on structures instituted in the fall of 1966, presumably to last until the end of 1967, was reported on average to have reduced 1967 capital investment plans by only 0.7 percent. In this case, where depreciation related only to structures, the equipment tax credit seemed more potent, inasmuch as the mean percentage reduction in investment plans as a consequence of its suspension was 2.2 percent.

In early 1965 the effective reduction in 1964 corporate tax rates of 3.8 percent, or 2 percentage points, was reported to have brought about an average increase of only 0.8 percent in 1964 investment, although profits before taxes were between two and three times investment. Thus each dollar of tax reduction was reported to bring in the neighborhood of some 20 cents in added capital expenditures. The proposed 10 percent corporate tax surcharge, raising the effective tax rate by some 4½ percentage points, was estimated in early 1968 to reduce capital spending plans for 1968 by a mean figure of 1.6 percent, implying again only about a 20 cent reduction in corporate capital expenditures for each dollar of increased corporate taxes.[6]

Both the accuracy and consistency of these responses is questionable. The bulk of individual survey answers as to percentage effect were "zero." There is no evidence that lack of information was translated into zero responses, but to the extent that it was, means would of course be biased toward zero and effects of tax measures underestimated. Running the other way, a number of relatively extreme observations, including figures of "+ 100 percent," contribute to higher means and make the differences between means of doubtful significance. Large percentage responses seem to come predominantly from small firms. We have consequently calculated ***weighted*** means of survey responses, and the estimated incentive effects, measured as total indicated changes in expenditures divided by total expenditures, are universally more moderate. Thus the anticipated total effect of depreciation guidelines in 1963 amounted to 1.9 percent of total expenditures, rather than the 4.2 percent which was the unweighted or simple mean of survey percent responses. For the equipment tax credit, a weighted mean of 1.4 percent for anticipated effects in 1963

---

[5] This despite estimates indicating that the present value of tax advantages due to the equipment tax credit was, for reasonable rates of discount, greater than that due to the depreciation guidelines, a fortiori, since it turned out that many firms had been "liberalizing" their depreciation before the new guidelines and hence had less additional gain. Hall and Jorgenson, pp. 50, 51, 55, estimate the impact on the rental price of capital as roughly one and one-half times greater for the equipment credit in 1963 and roughly two and one-half times greater in 1964, after firms were no longer required to deduct the credit from an asset's depreciable base.

[6] These figures are interestingly comparable with those of Coen (1971, p. 179). Dealing with manufacturing only, for the entire period 1954 to 1966-III, he estimated increased capital expenditures of $4.8 billion to be associated with total tax reductions of $13.7 billion, a ratio of some 35 percent.

TABLE 2—PERCENTAGE EFFECTS ON CAPITAL EXPENDITURES, SIMPLE AND WEIGHTED MEANS, VARYING AND IDENTICAL PANELS OF RESPONDENTS

				Identical Panels					
		All Observations Available		Anticipations, 1963 and 1964			Anticipated and Actual, 1964		
					Panel for Weighted Means			Panel for Weighted Means	
Expenditures	Tax Measure	Simple Means	Weighted Means	Simple Means	Simple Means	Weighted Means	Simple Means	Simple Means	Weighted Means
Anticipated for 1963	Depreciation Guidelines	4.2	1.9	4.4	2.8	1.2	—	—	—
	Equipment Tax Credit	2.8	1.4	2.6	1.2	0.7	—	—	—
Anticipated for 1964	Depreciation Guidelines	2.8	1.3	3.0	2.7	1.3	2.7	2.1	1.2
	Equipment Tax Credit	1.1	0.8	1.1	1.2	0.8	1.1	1.1	0.7
Actual, 1964	Depreciation Guidelines	1.5	0.9	—	—	—	1.5	1.5	1.0
	Equipment Tax Credit	1.3	1.1	—	—	—	1.5	1.5	1.4
Number of Respondents		[a]	[a]	217	178	178	206	182	182

[a] Varies as indicated in Table 1.

replaces the 2.8 percent unweighted mean.

Some inconsistency stems from the fact that samples were not identical from year to year (or for simple and weighted averages, because a number of observations did not include the actual dollar expenditures required in weighting). In particular, no downward trend in anticipated effects from 1963 to 1964 is apparent when we examine identical panels of respondents, as may be noted in Table 2.

It is possible to compare survey responses with *ex post* estimates derived from econometric estimates by Hall and Jorgenson, Bischoff, Klein and Taubman, and Coen. The survey responses generally indicate considerably less effectiveness of the various tax incentives, especially in terms of weighted means, than the predictions of Hall and Jorgenson. The 1964 expenditure effects viewed *ex post* come midway between those estimated by Klein and Taubman and by Coen and, for the depreciation guidelines, come quite close to those estimated by Bischoff. The largest discrepancies occur in estimates of effects of the equipment tax credit, where the econometric estimates, particularly those by Hall and Jorgenson, are much higher than the survey responses, and in the case of the short-run 1967 suspensions, where the econometric estimates are again higher. Only corporate income tax reductions, viewed as perverse or trivial in their results in two econometric estimates, appear to have greater, though still very small, effects in the surveys.

Examination by industry of the means of reported survey effects in Table 3 shows wide variations with particularly large figures reported for railroads and generally negligible ones for utilities, where the credit was only 3 percent and where indeed other evidence has indicated that capital expenditures are substantially dominated by demand and output considerations.

II. General Investment Equations Including the Tax Measure Survey Responses

The survey responses have been introduced as variables in regressions which also include the sales change, profits, and depreciation variables which were used earlier in explaining these capital expenditure data.[7] In such regressions, if a particular tax measure always had precisely the effect on capital expenditures indicated by respondents it would enter with a coefficient of unity. For example, if the respondent indicated that the new depreciation schedule increased the investment he planned for 1963 by 5 percent, an increase in planned investment by 5 percent over and above the

---

[7] The underlying investment model into which the tax policy response variables were fitted may be seen in Eisner (1967), particularly Tables 3 and 5, pp. 376 and 380.

TABLE 3—PERCENTAGE EFFECTS ON CAPITAL EXPENDITURES, SIMPLE MEANS OF SURVEY RESPONSES, BY INDUSTRY[a]

Expenditures	Tax Measure	Primary Metals	Metal Working	Chemical Processing	Other Manufacturing	Mining	Utilities	Petroleum	Railroads	Stores	Transportation and Communication Other Than Railroads
Anticipations for 1963	Depreciation Guidelines	2.5	4.3	5.6	5.6	11.7	0.2	0	9.8	1.0	2.3
		(19)	(99)	(42)	(58)	(18)	(28)	(9)	(12)	(25)	(11)
	Equipment Tax Credit	0.5	2.8	1.5	5.7	6.5	0.7	0	3.8	1.0	4.5
		(18)	(98)	(42)	(57)	(18)	(28)	(9)	(12)	(27)	(10)
Anticipations for 1964	Depreciation Guidelines	2.3	4.6	2.5	3.1	0.7	0.3	0	9.5	1.7	0.8
		(15)	(84)	(41)	(57)	(14)	(27)	(7)	(6)	(25)	(8)
	Equipment Tax Credit	1.1	1.7	1.1	1.1	0.7	0.6	0	1.3	0.7	0.2
		(15)	(84)	(41)	(57)	(14)	(25)	(7)	(6)	(25)	(8)
Actual, 1964	Corporate Income Tax Reductions	0.1	0.9	0.9	1.1	1.4	0.3	0.2	0.6	0.2	1.7
		(14)	(81)	(33)	(46)	(14)	(19)	(8)	(7)	(24)	(9)
	Depreciation Guidelines	0.9	1.1	1.3	2.2	0.7	0.4	0.1	13.4	1.7	2.2
		(14)	(81)	(33)	(45)	(14)	(18)	(8)	(7)	(24)	(9)
	Equipment Tax Credit	0.6	1.4	1.6	1.6	0.7	0.4	0.4	1.9	0.8	5.0
		(14)	(81)	(33)	(45)	(14)	(17)	(8)	(7)	(24)	(9)
Anticipations for 1967	Suspension of Accelerated Depreciation on Structures	−0.4	−2.7	−2.9	−5.2	0	−0.1	0	−1.9	0	0
		(19)	(100)	(41)	(55)	(9)	(30)	(9)	(13)	(26)	(10)
	Suspension of Equipment Tax Credit	−0.4	−1.4	0	−1.2	0	−0.0	0	0	0	0
		(19)	(100)	(41)	(55)	(9)	(30)	(9)	(13)	(26)	(10)
Anticipations for 1968	Corporate Income Tax Surcharge	−1.6	−2.5	−0.3	−4.0	0	0	0	−2.1	−0.5	0
		(19)	(82)	(33)	(42)	(14)	(23)	(6)	(12)	(22)	(10)

[a] Figures in parentheses are numbers of firms responding.

amount planned as a consequence of sales changes, profits, and depreciation charges, would imply a coefficient of unity for this response variable. If anticipated investment, despite the response, were actually only 2 percent more than indicated by the regression on the other variables, then the coefficient of the response variable would be 0.4. If the response variable in fact proved, despite respondents' claims, irrelevant to their anticipated expenditures, the coefficient should turn out to be zero. If ancipated expenditures actually moved counter to the direction indicated by the response, the coefficient would prove negative.[8]

Despite the availability of over 3600 observations, regression results were inconclusive. The depreciation guideline coefficients in firm time-series regressions, and less consistently in cross-sections and industry regressions, were roughly in the neighborhood of unity. But coefficients of the equipment tax credit variables were generally negative and of outlandish values in the time-series, while usually positive but varying substantially in cross-section regressions. Coefficients relating to the suspension of accelerated depreciation in 1966 were wildly negative (of the "wrong" sign as the variable was defined), while those of the suspension of the equipment tax credit were more often positive, but of considerable variance. Coefficients for the corporate income tax surcharge anticipated in 1967 and early 1968 were less than unity in firm time-series and cross-section regressions. In all cases standard errors were so high that parameter estimation and statistical inference were unreliable. There was, however, little in the results to suggest that the relatively moderate survey responses *underestimated* the effects of the tax measures, and the various negative coefficients raised further questions

---

[8] Algebraically, the regressions were of the form:

$$i = b_0 + \sum_{j=1}^{m} b_j x_j + \sum_{k=m+1}^{n} b_k x_k + u$$

where i = the ratio of capital expenditures, actual or anticipated, to gross fixed assets; x_j = the current and lagged relative sales change, profits and depreciation variables; and x_k = the tax policy survey response variables, expressed as ratios of actual or anticipated capital expenditures.

in some instances as to the positive effects reported.[9]

III. Conclusion

It is of course possible that survey respondents are unaware of all the influence, direct and indirect, that particular policies may have on investment. For what they are worth, however, survey responses have indicated only modest effects of tax measures designed to stimulate, or discourage, capital expenditures. Means of survey results were usually less than all except the smallest of estimates from several econometric models. Each dollar of loss or gain in taxes appeared to generate only very minor compensating capital expenditure. An attempt to estimate parameters of survey variables in general investment equations was inconclusive, but did not suggest that the survey responses, while of questionable accuracy, were underestimates.

APPENDIX

Basic data for econometric estimates in Table 1 are taken from Fromm as follows:

1. *Hall and Jorgenson*

Table 2-5 on pp. 46–47 lists gross investment divided into equipment and structures, by year:

	Equipment	Structures	Total
	(billions of 1965 dollars)		
1963	31.191	19.418	50.609
1964	35.116	20.604	55.720
1967 (est.)	39.408	24.433	63.841

Table 2-7, p. 52, lists estimated capital expenditures due to the guidelines and tax credit:

	Guidelines	Tax Credit
	(billions of 1965 dollars)	
1963	1.062	2.238
1964	1.220	3.765

Table 2-8, p. 54, reports estimated capital expenditures in 1964 due to the corporate tax cut, assuming no effect on the level of output, as −0.248 billions of 1965 dollars.

All of these incentive effects were divided by total gross investment to obtain the percentages shown in Table 1.

Table 2-11, p. 58, reports estimated effects on capital expenditures due to 1966 suspensions on the assumption that the suspensions would last through 1967:

	Equipment (tax credit)	Structures (accelerated depreciation)	Total
	(billions of 1965 dollars)		
1967	−4.062	−0.864	−4.926

Table 2-10, p. 56, reports estimated effects on capital expenditures associated with the actual abbreviated suspensions.

	Equipment	Structures	Total
	(billions of 1965 dollars)		
1967	−0.870	−0.289	−1.159

A figure for anticipated capital expenditure plans in the fall of 1966 when the suspensions were expected to last through 1967 was then obtained by subtracting from actual 1967 capital expenditures the difference between these two sets of effects. The percentages shown in Table 1 are the estimated full-term suspension effects divided by this reconstruction of anticipated capital expenditures:

$$\$63.841b - (\$4.926b - \$1.159b) = \$60.074b$$

2. *Bischoff*

Table 3-15, p. 117.

3. *Klein and Taubman*

Table 5-4, p. 238, offers gross investment estimates on the various assumptions: (1) no change in tax law; (2) changes in cash flow as a result of the equipment tax credit and the new guidelines; (3) changes in rates of return as well as cash flows:

---

[9] Detailed tabulations of some of the regression results involving observations from all of the surveys from 1955 through 1968 are available on request to the authors.

	(1) No. Change	(2) Change in cash flow only	(3) Change in rate of return and cash flow	(4) Ratio of incentive effect to investment $\frac{(2)-(1)}{(2)}$	$\frac{(3)-(1)}{(3)}$
	(billions of dollars)			(percent)	
1963	57.825	58.000	58.975	0.3	1.9
1964	63.350	63.925	66.400	0.9	4.6

Following Klein and Taubman's preference for the "cash flow only" estimate, the difference between "no change" and cash flow only, divided by the latter, is presented in our Table 1.

Table 5-2, p. 234, gives estimates of 1967 investment under the assumption of no suspension of incentives and that of a temporary suspension lasting through 1967.

No Suspension	Temporary Suspension
(billions of dollars)	
84.475	82.175

The difference in estimates divided by the temporary suspension figures yields the percentage effect.

4. *Coen*

Table 4-12, p. 178, reports actual investment in manufacturing and estimates of investment attributable to the whole package of tax incentives including the accelerated depreciation dating from 1954. Coen offers several sets of estimates. For his preferred variable-adjustment-speed model with a twelve-quarter inverted-V lag and user cost based on declining balance depreciation, the relevant figures are:

	Attributable to Incentives	Manufacturing Investment	Ratio of Incentive Effects to Investment
	(millions of 1964 dollars)		(percent)
1963	422	12,673	3.3
1964	704	14,848	4.7

Our estimate for 1963 and 1964 incentive effects exclusive of the 1954 acceleration of depreciation was obtained by subtracting the $179 million of investment Coen finds attributable to incentives for 1961, when this depreciation effect was last isolated. The figure had been decreasing according to Coen's results. Thus later estimates listed for Coen, entailing the $179 million subtraction, are probably a bit too low.

REFERENCES

C. W. Bischoff, "The Effect of Alternative Lag Distributions," in G. Fromm, ed., *Tax Incentives and Capital Spending*, Washington 1971, 61–130.

R. M. Coen, "Tax Policy and Investment Behavior: Comment," *Amer. Econ. Rev.*, June 1969, *59*, 370–79.

———, "The Effect of Cash Flow on the Speed of Adjustment," in G. Fromm, ed., *Tax Incentives and Capital Spending*, Washington 1971, 131–96.

R. Eisner, "A Permanent Income Theory for Investment: Some Empirical Explorations," *Amer. Econ. Rev.*, June 1967, *57*, 363–90.

———, "Tax Policy and Investment Behavior: Comment," *Amer. Econ. Rev.*, June 1969, *59*, 379–88.

G. Fromm, *Tax Incentives and Capital Spending*, Washington 1971.

R. E. Hall and D. W. Jorgenson, Tax Policy and Investment Behavior," *Amer. Econ. Rev.*, June 1967, *57*, 391–414.

——— and ———, "Application of the Theory of Optimal Capital Accumulation," in G. Fromm, ed., *Tax Incentives and Capital Spending*, Washington 1971, 9–60.

D. W. Jorgenson, "Capital Theory and Investment Behavior," *Amer. Econ. Rev. Proc.*, May 1963, *53*, 247–59.

L. R. Klein and P. Taubman, "Estimating Effects within a Complete Econometric Model," in G. Fromm, ed., *Tax Incentives and Capital Spending*, Washington 1971, 197–242.

The U.S. Phillips Curve and International Unemployment Rate Differentials: Comment

By NANCY SMITH BARRETT*

A recent article in this *Review* by Robert Flanagan on international differences in unemployment experience draws some rather strong conclusions about the nature of unemployment flows in the United States relative to countries of Western Europe, as well as the relative efficiency of the *U.S.* labor market. However, since Flanagan uses a definitionally incorrect measure of one of the most important variables in his study, the interpretation of his results is at best questionable. In addition, there are data problems that Flanagan fails to address.

Using a different data source and correcting the error in the variable definition, my findings support some of Flanagan's theoretical conclusions and policy suggestions. However, there are some important differences that affect both the interpretation of the results as well as their policy implications.

I. Synopsis

Two major points stand out in Flanagan's paper. First, he contends that high unemployment rates in the United States relative to those in Western Europe (with Sweden and Great Britain used as models) reflect real differences in the tradeoff between inflation and unemployment. That is, the lower European unemployment rates are not due to a willingness to accept more inflation or to differences in statistical concepts and measurement, but signify a different and more favorable Phillips relationship.

Second, Flanagan attributes international differences in the Phillips curves to differences in the flow of workers into unemployment rather than to differences in the duration of their unemployment spells. He also suggests that these findings reflect a lower level of labor market efficiency in the United States relative to the European countries.

My analysis of the Swedish labor market compared with that of the United States supports the first proposition, but not the second. Since I used different data from Flanagan, the fact of our agreement on the first point lends independent support for some of his policy recommendations for improving the *U.S.* inflation-unemployment tradeoff. On the other hand, although I agree that there are real international differences in the Phillips tradeoff and that these are related to the components of unemployment he describes, I have found that differences in the flow into new unemployment between the United States and Sweden are substantially greater than differences in their overall unemployment rates. This implies (from Flanagan's equation (2)) that the average duration of unemployment is longer in Sweden than in the United States. As we shall see, this has important implications for the interpretation of the first finding in terms of the relative efficiency of the *U.S.* labor market as well as for stabilization policy.

II. Methodological Considerations and Data Problems

As Flanagan points out, the unemployment rate is the product of two components

$$UN = I_u D_u \tag{1}$$

where I_u is the weekly flow of individuals into new unemployment as a percent of the labor force, and D_u is the average duration of a spell of unemployment.

From a theoretical perspective, this breakdown is extremely useful for comparing the unemployment experience of different groups. A 4 percent rate of unemployment, for example, could represent a .80 percent weekly flow of individuals into new unemployment with five weeks average duration, or a .33 percent weekly flow and twelve weeks average duration. In the first case, workers become unemployed more often, but find new jobs rather quickly. In the second case, there is a greater tendency to hold on to jobs or to stay out of the labor market, but workers take more than twice as long to find a job once they become unemployed. As Flanagan observed, the determinants of I_u and D_u are different. This has implications for policies to reduce unemployment, for the welfare of the workers experiencing unemployment, as well as for the efficiency of the labor market.

---

* Professor of economics, American University. Curt Canarp and Christer Wretborn of the Swedish Labor Market Board (*AMS*) provided unpublished data on unemployment flows in Sweden.

Although it is interesting to compare the components of unemployment for different labor markets, there are complicated theoretical and technical problems associated with measuring I_u and D_u from unemployment data. Most of these stem from the ambiguous nature of the unemployment concept, particularly when applied to changes in very short periods of time. Flanagan does not discuss these difficulties at all, and the casual reader is likely to assume that his measure of I_u (he never (correctly) analyzes D_u) is as reliable as most other unemployment statistics. Using a different (and from my perspective, better) data source, I obtained different results from Flanagan. Because of this, it will be necessary to explore the data problems before discussing the data.

In the United States and Sweden[1] the official measure of unemployment is based on monthly rotating samples in which individuals are asked if they were employed, unemployed, or not in the labor force during the survey week. Those unemployed are also asked how long they have been in that status. Flanagan uses data from these monthly surveys to measure I_u, the number of persons unemployed less than one week (as a percent of the labor force). There are two major problems with this. First, because an unemployed worker must be actively seeking work, there is some technical movement in and out of the labor force from week to week for persons who want to work but temporarily stop looking. This behavior is particularly likely for women and teenagers whose association with the labor force is somewhat more tenuous than for adult males. A preferable measure of I_u would ignore these very short-run changes in labor market status.

Second, Daniel Suits and Richard Morgenstern, Hyman Kaitz (1970, 1972), and others have found that some persons experiencing one or two weeks of unemployment (exactly those represented by I_u) do not show up in the monthly surveys. This is because a person who is laid off Monday morning is not counted as unemployed during the week, although someone who entered the labor force Monday morning is. Typically, the monthly surveys show more persons with two weeks unemployment than one week. The problem cannot be solved by looking at changes in successive weeks either (as Flanagan presumably did), because more people report unemployment of 4, 8, 12, . . . weeks than is consistent with a normal pattern of unemployment survival rates. This bunching of the data around weeks that are multiples of four suggests that respondents tend to report unemployment duration in months rather than weeks.[2]

Instead of using data from the monthly survey, I have estimated I_u from the annual labor force surveys conducted in each country that ask individuals how many weeks they were unemployed and how many spells of unemployment they experienced over the previous year.[3] I_u can be computed from

---

[1] Since the United States and Sweden were polar cases in Flanagan's study, I omitted analysis of Great Britain, the intermediate case.

[2] A referee has pointed out that Flanagan's actual estimate of I_u for the United States appears to be biased downward relative to other estimates from the monthly surveys. He estimates new entrants to unemployment each week at about 470,000 in the United States, p. 121, as compared with the Kaitz (1972) estimate of 550,000 to 750,000 for the same period. This means that Flanagan's estimate of average duration for the United States is too high.

[3] In the United States, the monthly surveys are the

$$I_u = \frac{N \times S}{52\, L} \tag{2}$$

where N is the number of individuals experiencing unemployment during the year, S is the average number of spells per person, and L is the average labor force. Although the annual surveys contain some biases from retrospective reporting (see Morgenstern and the author), they are the only source of data for the experience of workers over an entire year.[4]

In both countries, failure to look for work for less than three weeks is not considered an interruption of an unemployment spell, although those weeks are not counted in the computation of average duration. Thus, we would expect our estimate of I_u to be somewhat less than Flanagan's, with the greatest discrepancy for women and young people.

The second component of the unemployment rate is D_u, the average length of the unemployment spell. In addition to the data problems I have discussed, Flanagan uses a definitionally incorrect measure of D_u that is biased in a predictable way. To test the hypothesis that D_u does not vary between countries, Flanagan uses the rate of long duration unemployment, that is, the percent of the labor force unemployed fifteen weeks or more. This measure of D_u tells us nothing about the average duration per spell when there are differences in unemployment flows between the two countries. In fact, if the high duration unemployment rates are the same for two countries, then the one with a higher flow into long duration unemployment will have shorter durations per spell.[5] Although Flanagan purports to determine whether the "average time spent between jobs" (p. 127) varies between countries, his conclusion that international differences in duration are not significant (based on a comparision of long duration unemployment rates) is not accurate when a correct measure of D_u is used.

There are some problems in measuring D_u. The duration concept required by (1) is the average length of the completed unemployment spell, not the length of time the unemployed have been unemployed.[6] Since the duration measured in the monthly surveys is average age rather than life expectancy, there may be some bias in using these data for D_u. Consequently, I computed D_u from the annual surveys for the United States.[7] Since the annual surveys were only available for three years in Sweden, I used the monthly surveys to estimate D_u after ascertaining that there was very little difference between the two measures in the overlapping years.[8]

Current Population Survey (*CPS*) and the annual survey is the *Work Experience Survey* (*WES*). In Sweden, the monthly survey is the *Arbetskraftsundersökningen* (*AKU*) and the annual survey, compiled by the Swedish Labor Market Board (*AMS*) is generally referred to as the *AMS* survey. Since the *AMS* survey is only available for three years, we extrapolated from the Swedish monthly surveys on the basis of relationships found for those three years. This will be discussed below.

[4] There are two major types of problems associated with the annual surveys. The first are retrospective reporting biases. Morgenstern and the author have shown these biases are insignificant for central-age white males and most important for women, blacks, and teenagers. This does not affect our criticism of Flanagan's results since our conclusion holds for central-age white males as well as for other demographic groups. A second type of problem with the annual surveys is related to some technical points such as the truncation of unemployment spells at the beginning and end of the year. These problems, however, should not bias international comparisons when the surveys being compared deal with these technicalities similarly.

[5] Long duration unemployment does not measure the average duration of all unemployment spells, either. But for Sweden and the United States, differences in D_u are reflected in corresponding differences in long duration unemployment.

[6] The required measure of duration is analogous to life expectancy while the monthly surveys measure average age of unemployment. Unlike chronological age where life expectancy declines with age, average age of unemployment may be greater than life expectancy to the extent that the expected length of the unemployment spell increases with the survival rate. For a discussion, see Suits and Morgenstern.

[7] Although there is a slight downward bias associated with those individuals who were unemployed at the end of the year (and had not completed their last spell of unemployment) this measure is closer to the life expectancy concept required in equation (1) than the monthly survey data.

[8] This suggests that in Sweden the expected length of the unemployment spell does not vary very much with the survival rate.

TABLE 1—THE WEEKLY FLOW INTO NEW UNEMPLOYMENT (I_u) AND AVERAGE DURATION PER SPELL (D_u) (IN WEEKS) IN SWEDEN AND THE UNITED STATES

	Male				Female			
	I_u	D_u	UN_P	UN_A	I_u	D_u	UN_P	UN_A
			United States[a]					
1964	.518	8.4	4.4	4.1	.542	10.3	5.6	5.5
1965	.452	7.5	3.4	3.6	.449	9.4	4.2	5.0
1966	.391	6.6	2.6	2.8	.433	8.9	3.9	4.3
1967	.392	6.6	2.6	2.7	.424	8.9	3.8	4.6
1968	.352	6.5	2.3	2.6	.419	8.8	3.7	4.3
1969	.381	6.6	2.5	2.5	.402	8.4	3.4	4.2
1970	.488	8.5	4.1	4.0	.478	10.2	4.9	5.4
Average	.425	7.2	3.1	3.2	.450	9.3	4.2	4.8
			Sweden					
1966	.167	7.8		1.3	.263	7.6		2.0
1967	.177	11.3		2.0	.202	11.4		2.3
1968	.168	13.7		2.3	.173	12.1		2.1
1969	.131	13.5		1.8	.200	10.0		2.0
1970	.101	12.9		1.3	.164	11.9		1.8
1971	.193	12.4		2.4	.181	14.9		2.7
1972	.165	15.1		2.5	.171	17.5		3.0
Average	.157	12.4		1.9	.193	12.2		2.3

Sources:

Sweden: D_u and UN_A, Statistika Centralbyrån, *Arbetskraftsundersökningen* (*AKU*); I_u computed from $UN_A = I_u D_u$.

United States: D_u and I_u computed from data from Bureau of Labor Statistics, *Work Experience Survey of the Population* (*WES*). Based on computed survival rates of unemployment, for men the point estimate for those in the category "27 or more weeks of unemployment" is 33 weeks and for women it is 44. UN_P computed from $UN_P = I_u D_u$; UN_A, Bureau of Labor Statistics, *Current Population Survey* (*CPS*).

[a] (Whites only.) The data for 1964 and 1965 include 14- and 15-year-olds who are excluded thereafter.

III. Empirical Evidence

Table 1 shows D_u and I_u for Sweden and the United States for those years for which data are available. For the United States, both D_u and I_u are computed from the retrospective survey,[9] UN_P is the unemployment rate obtained from equation (1) and UN_A is the official (*CPS*) unemployment rate. The difference between UN_A and UN_P is the bias from retrospective reporting.

In Sweden, on the other hand, the retrospective surveys were available for only three years. Furthermore, they tended to overstate the actual amount of unemployment. Most likely this was due to the somewhat ambiguous nature of the unemployment concept in Sweden where the unemployed are often provided temporary, relief-type, government jobs that in retrospect are recalled as "unemployment." Nevertheless, the estimate of duration from the annual survey was sufficiently close to that in the monthly surveys (an average of 13.85 weeks

---

[9] The relation between my estimate of I_u and Flanagan's is interesting. For 1970, the only year for which he provides data, his I_u for white males is .42 (relative to my .49) and for white females, .70 (relative to my .48). This suggests that the basic problem with his measure is associated with the short-run changes in labor market status that characterize female workers. This is consistent with our finding in an earlier paper that the greatest discrepancy between the monthly and annual unemployment surveys is for female workers. Possible explanations are discussed in the paper of the author and Morgenstern.

TABLE 2—PERCENT DISTRIBUTION OF DURATION OF UNEMPLOYMENT IN SWEDEN (1970) AND THE UNITED STATES (1969)

Sweden	less than 5 weeks	5–8	9–12	13–26	27+	
Percent	49.9	17.6	8.8	13.0	10.7	
Cumulative percent	49.9	67.5	76.3	89.3	100.0	
United States[a]	less than 5 weeks	5–6	7–10	11–14	15–26	27+
Percent	57.5	9.3	12.9	7.1	8.5	4.7
Cumulative percent	57.5	66.8	79.7	86.8	95.3	100.0

Sources: Sweden: *AKU*, United States: *CPS*.

[a] Includes blacks.

compared with 12.83 weeks in the monthly surveys) that I felt justified in using the duration estimates from the monthly surveys for D_u.[10] If anything, the monthly surveys may understate D_u somewhat. I computed I_u implicitly from the official unemployment rate, using equation (1).

Flanagan contends that differences in the *U.S.* and Swedish unemployment rates are due entirely to differences in I_u. But these data show that although Swedish unemployment rates are about one-half those in the United States, Swedes experience significantly longer average spells of unemployment than Americans. Swedish males have unemployment spells that are nearly twice as long as white males in the United States.[11] It appears that Flanagan underestimated I_u for the United States relative to Sweden. Furthermore, because he never estimated average duration per spell as required by equation (1), he failed to observe the sizeable differences in D_u between the two countries.

Because of Flanagan's interest in long-term unemployment, I also examined the distribution of duration among the unemployed. Table 2 shows the percent distribution of duration of unemployment in Sweden and the United States for the most recent high employment year in each country. Over 10 percent of the Swedish unemployment was over twelve weeks in duration while less than 5 percent of the *U.S.* unemployment was for more than fourteen weeks. It appears that Flanagan found no difference in the long duration unemployment rates (U_D) for the two countries because U_D contains an element of turnover which is higher for the United States. That is, since the United States has higher unemployment rates, long duration unemployment represents a smaller proportion of total unemployment. In fact, Flanagan himself estimates that if both Sweden and the United States had the same unemployment rate, the long duration unemployment rate would be higher in Sweden. Because his sample contained earlier years than mine, however, there may have actually been somewhat less intercountry variation in D_u than I found since there is evidence of a secular upward trend in duration for Sweden but not for the United States.[12]

IV. Implications

Economists are concerned about unemployment for a number of reasons. Flanagan seems to view the problem as one of labor market maladjustment; that is, a failure to achieve equilibrium. Others are concerned with the welfare of the individuals experiencing unemployment and infer from international differences in unemployment rates that workers are worse off in the United States. In addition, there is the concern for the foregone output that is seen as the opportunity cost of price stability. Thus, the opportunity cost of stopping inflation is often considered to be higher in the United States than in Western Europe.

---

[10] The monthly survey estimates were used even where the annual survey estimates were available for consistency.

[11] Blacks were deliberately excluded from the *U.S.* data so that our results could not be attributable to demographic factors.

[12] There seems to be an error in Flanagan's Table 5 in which the coefficient on the time trend in the Swedish equation is insignificant, contrary to the textual exposition.

Our finding that Swedish workers become unemployed less often than Americans but take longer to find jobs once they become unemployed has important implications for all these concerns. First, if labor market efficiency is related to the *speed* with which job seekers are matched with vacancies, the *U.S.* labor market is more efficient in this respect than the Swedish. An argument could be made that efficiency is not reflected in few job changes but in the ease with which job changes can be made. Thus, Flanagan's contention that the *U.S.* labor market exhibits the most incomplete adjustment of supply and demand does not seem to be the case.

With respect to the welfare of the worker who experiences unemployment, the analysis is complicated by an inability to distinguish between voluntary and involuntary unemployment. Within a search theoretic model, high I_u may reflect more efficient (and hence beneficial) sampling of the job market or it may imply excessive involuntary separation rates. Robert Hall has suggested on the basis of demographic analysis that the major unemployment problem in the United States is not long periods of chronic unemployment but frequent job changes and unemployed search by individuals who are dissatisfied with their jobs or cannot advance their careers. Blacks, unskilled workers, and other groups in lower labor market strata flow more frequently into unemployment than central-age white males. This interpretation suggests that the question of worker welfare must be addressed in the context of employment experience as well as the unemployment experience.

Low I_u is not necessarily indicative of more efficient labor markets or higher levels of worker satisfaction when duration of unemployment is expected to be long. I have argued in another connection (with Bo Södersten) that the longer durations of unemployment in Sweden represent a greater expected search cost for unemployed individuals. This may help to explain why I_u is so low. Swedish workers may be more reluctant than Americans to change jobs (or at least to undergo unemployed search) and will resist threatened layoffs more strongly. If workers become locked into unsatisfactory jobs because of labor market inefficiency—that is, the failure to match seekers with jobs in a reasonable period of time—this *captive worker effect* may understate the actual welfare loss associated with a given unemployment rate. Thus, although Swedish workers have lower unemployment rates they are not necessarily better off than American workers. In fact, Flanagan observes that the relatively high I_u in the United States may imply that the cost of search is sufficiently low that a larger fraction of the labor force finds unemployed search rational. I argue that this is indeed likely to be the case relative to Sweden.

Other findings of the Flanagan study lend support to the hypothesized captive worker effect. He found, for instance, that there is a strong inverse relationship between the separation rate and the unemployment rate in Sweden while no such relation exists for the United States. This implies that in slack labor markets Swedish workers remain in their jobs or drop out of the labor force, while this is not the case in the United States.[13]

Flanagan notes that emphasis in *U.S.* labor market policy has been on reducing the length of the unemployment spell. In Sweden, on the other hand, attempts to reduce unemployment have focused on reducing the number of persons flowing into unemployment. These policy preferences are undoubtedly reflected in the components of unemployment experience in the two countries. But this does not imply that the captive worker effect is not operative in Sweden. In fact, it may be a manifestation of the effect. Workers are prevented from experiencing unemployment by being placed in a state program that provides low-status, low-paying, relief-type jobs. The effect of this policy is to lower the unemployment rate, but the welfare implications are far from clear.

---

[13] This is consistent with Flanagan's finding, p. 127, that labor force elasticity is higher in Sweden than in the United States.

Analysis of the demographic profile of unemployment flows in Sweden provides more evidence of the captive worker effect. In Figure 1, D_u and I_u are plotted by age.

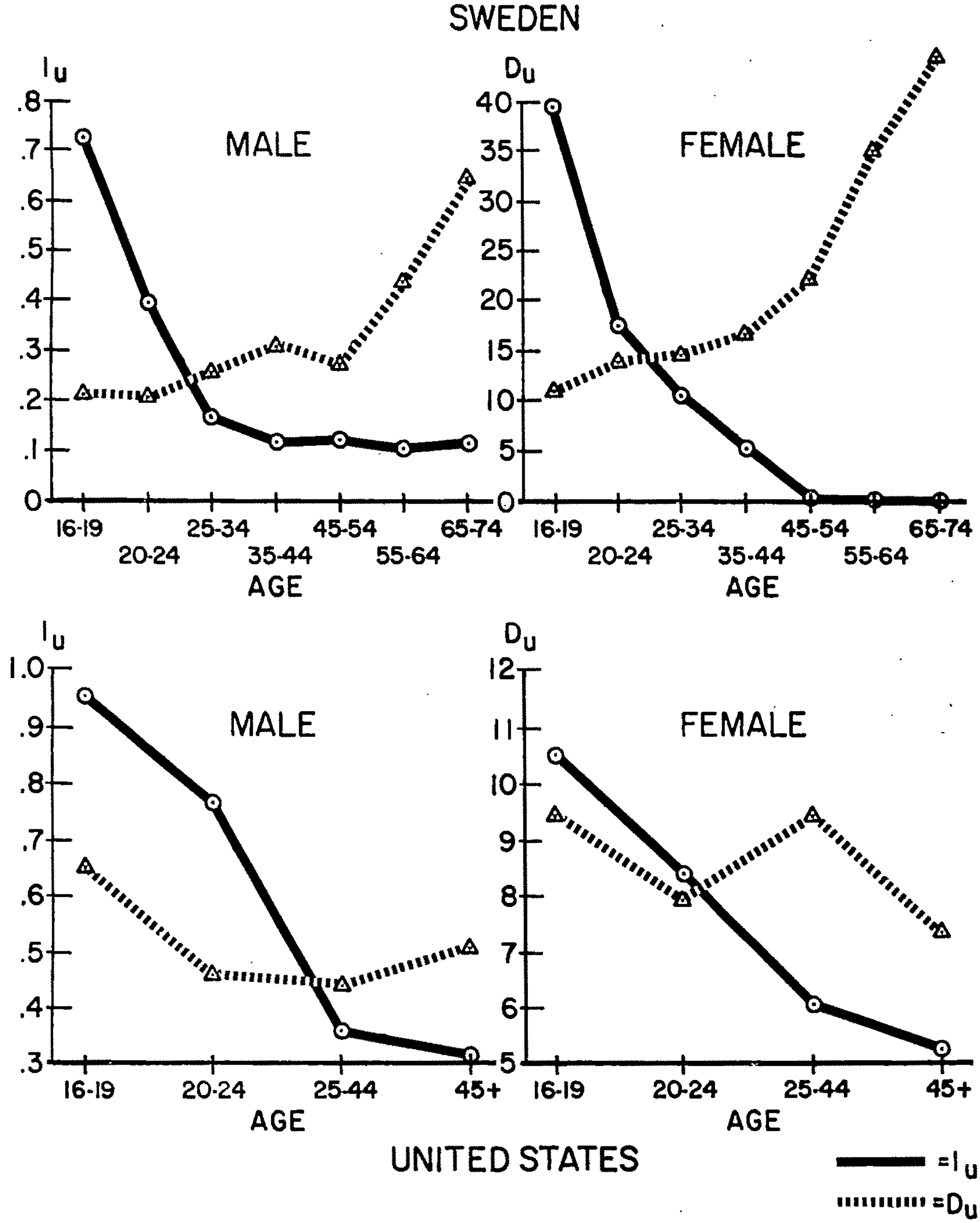


FIGURE 1

For Sweden, those age groups with long average durations of unemployment have much lower flows into new unemployment than workers who have less trouble finding jobs.[14] In the United States, on the other hand, there is no clear inverse relationship between D_u and I_u across age groups, suggesting that since expected duration is relatively short for all groups, it has less effect on workers' willingness to undergo unemployed job search. In the United States, variations in D_u and I_u seem to be more closely related to sociological factors than to labor market efficiency. (For a discussion see the author and Morgenstern.)

Flanagan considers the possibility that in tight labor markets a decline in D_u might induce a greater than proportional increase in I_u, causing the unemployment rate to rise. Using only the American data, he concludes that this type of behavior is unlikely. However, in Sweden, to the extent that workers have historically been discouraged from unemployed job search because of the expectation of a long spell of unemployment, such behavior is more likely to occur. In 1971, for instance, the unemployment rate for men nearly doubled, although the average duration actually declined. In such a case, the interpretation of the unemployment rate as a measure of labor market tightness is questionable. In the high unemployment year it took less time to find a job than in the "tighter" labor market of 1970. Although unemployment rates in Sweden over the period 1966–70 remained relatively stable, there was a tendency for duration of unemployment to rise with a concomitant decline in mobility. This suggests that Swedish workers may have experienced a welfare loss over the period that was not apparent from the raw unemployment data. The sizeable jump in the unemployment rate in 1971 and 1972 may represent a more apparent fruition of a structural problem hidden in the captive worker effect rather than an overall slackening of the labor market.

My suggestion that the welfare loss associated with unemployment might be partly related to the employment experience also has implications for the Keynesian concern with unemployment in terms of foregone output. Search theorists argue that unemployment may be productive if unemployed search leads to job changes that increase worker productivity. The captive worker effect may prevent such productivity increases. On the other hand, frequent job changes and flows into unemployment such as characterize the American economy may also be unproductive, particularly where on-the-job training is not transferable. However, there clearly is not a one-to-one correspondence between the unemployment rate and foregone output. That is, just because the *U.S.* unemployment rate is twice as high as in Sweden, this does not mean there is a greater waste of labor potential.

V. Policy Considerations

Our results support Flanagan's view that the way in which the unemployed experience their unemployment varies markedly between Sweden and the United States. In the United States the problem is not finding a job, but keeping the one presently held. The reverse holds in Sweden. In both cases, the welfare effects may be more closely related to the employment experience rather than the unemployment experience. Increased job satisfaction may cause a reduction in job changes and the flow into unemployment (I_u) in the United States. In Sweden, an increase in unemployment need not be viewed with alarm if duration is falling and more workers are moving to better jobs.

The most important conclusion to be drawn is that the unemployment rate is an imperfect measure of the welfare loss associated with unemployment, and that it is not necessarily indicative of the degree of structural maladjustment in the labor market. This is true not only when making international comparisons, but also when evaluating substrata within a national labor market. The notion that there is some full employment rate below which aggregate demand remedies are ineffective is also questionable, since the same unemployment rate may be associated with different patterns of unemployment flows. Although the Keynesian legacy has accustomed us to view the unemployment rate as the appropriate policy target, it does not recognize important variations in labor market flows that are the key to effective labor market policy.

---

[14] The same inverse relationship between D_u and I_u is evident in the regional data for Sweden. See *Årsarbetslöshetens struktur 1970, AMS*. The relationship is not as systematic in the *U.S.* regional data.

REFERENCES

N. S. Barrett and R. D. Morgenstern, "Why Do Blacks and Women Have High Unemployment Rates?" *J. Hum. Resources*, Fall 1974, *9*, 452–64.

——— **and B. Södersten**, "Unemployment Flows, Welfare, and Labor Market Efficiency in Sweden and the United States," *Swedish Econ. J.*, Sept. 1975, *77*.

R. J. Flanagan, "The U.S. Phillips Curve and International Unemployment Rate Differentials," *Amer. Econ. Rev.*, Mar. 1973, *63*, 114–31.

R. E. Hall, "Why is the Unemployment Rate So High at Full Employment?" *Brookings Papers*, Washington 1970, *3*, 369–402.

H. B. Kaitz, "Analyzing the Length of Spells of Unemployment," *Mon. Lab. Rev.*, Nov. 1970, *93*, 11–20.

———, "The Duration of Unemployment," *Proc. Amer. Statist. Assn. Soc. Statist. Sec.*, Washington 1974, 312–16.

R. D. Morgenstern and N. S. Barrett, "The Retrospective Bias in Unemployment Reporting by Sex, Race, and Age," *J. Amer. Statist. Assn.*, June 1974, *69*, 355–57.

G. L. Perry, "Changing Labor Markets and Inflation," *Brookings Papers*, Washington 1970, *3*, 411–41.

R. E. Smith and C. C. Holt, "A Job Search–Turnover Analysis of the Black-White Unemployment Ratio," *Ind. Relat. Res. Assn, Proceedings of the Twenty-Third Annual Winter Meeting*, Madison 1971, 76–86.

D. B. Suits and R. D. Morgenstern, "Duration as a Dimension of Unemployment," paper presented at the meeting of the Econometric Society, Winter 1967.

Arbetsmarknadsstyrelsen (*AMS*), *Årsarbetslöshetens strukter 1970*, *4*, Stockholm 1972.

———, Unpublished data from the annual surveys in the *Arbetskraftsundersökningen* (*AKU*), 1967–70, Stockholm 1973.

Statistika Centralbyrån (*SCB*), *Arbetskraftsundersökningen* (*AKU*), 1965–72, monthly surveys, Stockholm.

U.S. Bureau of Labor Statistics, *Current Population Survey* (*CPS*), data reported in *Manpower Report of the President*, various annual issues, Washington.

———, *Work Experience Survey of the Population* (*WES*), Special Labor Force Reports for 1964–70, Washington.

The U.S. Phillips Curve and International Unemployment Rate Differentials: Reply

By ROBERT J. FLANAGAN*

The opening paragraphs of the comment by Nancy Barrett promise far more disagreement than the rest of the paper achieves. It may minimize reader confusion if the concepts at issue are briefly reviewed before evaluating the specifics of Barrett's comment. In my paper I advanced the definition

$$U \equiv I_u \cdot D_u \tag{1}$$

as a more useful frame of reference for organizing discussions of unemployment than alternative decompositions. ($I_u = I/L$ is the proportion of the labor force becoming unemployed each week and D_u is the average duration of unemployment.) There is no disagreement between us on that point. Instead, the substantive analysis in Barrett's comment raises two conceptual issues: (a) whether published labor force survey data generally correspond to the definitions of U, I_v, and D_u; and (b) the validity of strict applications of steady-state equilibrium relationship (1) in a decidedly disequilibrium world. I shall briefly discuss each of these issues and then note their implications for the claims made in the comment.

Labor force survey data on unemployment by duration do not generally correspond to the completed duration of unemployment concept implicit in equation (1). Staying with the steady-state world for the moment, I individuals become unemployed every week as a result of quit, layoff, and labor force entry behavior. Some fraction $(1-p)$ of these individuals will complete their spell of unemployment by accepting employment or withdrawing from the labor force during their first week of unemployment, and pI $(0<p<1)$ of the cohort go on to experience a second week of unemployment. If the continuation rate p remains constant as a given weekly cohort of the unemployed ages, p^2I will experience a third week of unemployment, p^3I a fourth week, and so on. Under these assumptions it can be shown that the average duration of completed unemployment is

$$D_u = \frac{1}{1-p} \tag{2}$$

Thus, under these assumptions, Barrett's comment that D_u is higher in Sweden implies that the unemployment continuation rate p is higher in Sweden than in the United States.

If the continuation rate is constant over a spell of unemployment, the distribution of unemployment by duration reported in labor force surveys should be the same as the distribution of completed duration. However, in an important article, Hyman Kaitz demonstrated that if the continuation rate rises with the duration of a spell of unemployment, the labor force survey data will overstate the completed duration of unemployment (while a declining continuation rate will have the opposite effect). Much of the earlier discussion in the Barrett comment revolves around the argument that the gap between completed duration and reported duration in the survey data is greater in the United States than Sweden, so that the difference in I_u between the countries was understated. This reduces to an assertion that unemployment continuation rates rise more rapidly in Sweden than in the United States.

The repeated reapplication of equation (1) in Barrett's comment may lead some readers to overlook the fact that the focus of my study was a problem of disequilibrium behavior—labor market factors contributing to international differences in Phillips curves. As such, most of the empirical analysis of that paper emphasized disequilibrium (cyclical) behavior of the incidence of unemployment, separation rates, and long duration unemployment. This raises questions of the applicability of the steady-state equilibrium relation (1) in labor markets which are subject to cyclical shocks. The question may be stated as follows: During a cyclical contraction (expansion) will data on I_u and D_u yield accurate measures (predictions) of U when substituted into equilibrium relation (1) (as, for example, Barrett does in computations in Table 1)?

---

* Assistant professor of labor economics, Graduate School of Business, University of Chicago.

Consider two changes which would generate increased unemployment during a recession. In scenario number 1, I_u, the weekly flow of unemployment remains unchanged, but the probability of incurring another week of unemployment p rises. Suppose initially that $I_u=.01$ per week and $p=.50$, then by equations (1) and (2), in equilibrium completed duration $D_u=2$ weeks and $U=.02$. Now suppose that with weakened demand, p rises to .75. With $I_u=.01$, we now have $D_u=4$ and $U=.04$ *in equilibrium*. However, this equilibrium is not achieved until the full effects of the increase in p work their way through the system many weeks later. During the period of adjustment, however, actual unemployment is considerably below predicted unemployment because the number of unemployed in each duration category is dominated for a while by the old (lower) continuation rate. In scenario number 2, the recession is characterized by a rise in the weekly incidence (say $I_u=.02$) but the continuation rate remains unchanged (say $p=.5$). Originally $U=.02$. With $I_u=.02$, the new equilibrium unemployment rate will be $U=.04$. In the first week, however, there is only one cohort of $I_u=.02$ (the most recent cohort) searching for a job so that the actual unemployment rate is approximately .03 rather than .04, a rate that will not be reached for several weeks.

Both of these examples carry the same message: during a period in which I_u or p is increasing, the equilibrium relation (1) will *overstate* the actual amount of unemployment in the economy. In an expansion, the converse is true: the equilibrium relation will *understate* the true amount of unemployment.

With the foregoing as background, I turn to an evaluation of Barrett's comment. There appear to be three issues of substance which I shall discuss in turn: 1) the appropriate empirical definition of I_u; 2) my analysis of the duration of unemployment; and 3) the welfare implications of a given pattern of empirical results on unemployment.

In her analysis of I_u, Barrett argues that the difference in I_u between the United States and Sweden is even greater than the results of my article indicate. After reviewing Barrett's empirical analysis along with information from other sources, I find this congenial possibility resistable. On the issue of the appropriate empirical definition of I_u, Barrett mistakenly assumes I use data from monthly surveys to measure I_u and thus pick up some technical involvement in and out of the labor force. In fact, the Swedish I_u statistic was estimated from annual data and is defined as the annual average number of people unemployed one week or less as a percent of the labor force. In any event, this should not be a major source of difference, since the spells data, which she favors, should include short spells as well.

The use of spells data will circumvent the "bunching" problem, but Barrett's analysis is flawed in two important respects. First, after arguing the merits of the unemployment spells data to motivate her equation (2), she does not use spells data for her computations of the Swedish I_u in Table 1. Instead, data on U and D_u are used to compute I_u using the steady-state equilibrium relation (1). As demonstrated above, this application of equilibrium concepts to a disequilibrium world yields biased results.

The second problem arises from Barrett's application of unemployment spells data in the *U.S.* computations in Table 1. The use of these data requires applying an arbitrary judgment when the number of spells is high. In the source cited by Barrett, *U.S.* data on unemployment spells is truncated at the category 3+ spells. To develop an estimate of S, the average spells per person in Barrett's equation (2), one must make an assumption about the distribution of actual spells across people in the 3+ category. This ***assumption is crucial*** to the magnitude of the difference between the I_u measured by the reported duration of survey respondents on the one hand, and by reported spells on the other, and this alleged difference is at the core of Barrett's argument about the relative size of I_u and D_u. Yet the text of the comment offers no statement of the assumption used and why it is justified. If, for example, Barrett simply applied three or four spells (or some similar constant to all people in the 3+ category, then the differences between her numbers and those reported in my article would be greatest for those groups experiencing the most frequent labor force turnover. This is in fact what she reports in footnote 9, but to the extent that it is an artifact of a spells assumption this difference is not real. Until the nature of the arbitrary assumption regarding *U.S.* spells is clarified and until spells data are actually used in the Swedish computations, the magnitude of the differences reported by Barrett must be regarded as suspect.

The second part of Barrett's comment concerns the duration of unemployment. There are two important reasons why the average duration of unemployment is not analyzed directly in my article. First, D_u is an extremely ambiguous dependent variable. Theories of optimal search behavior often yield implications for this statistic, but these are best tested with microeconomic data. On the other hand, D_u may also be sensitive to structural barriers in labor markets, and the use of long duration unemployment data to examine this possibility is conceptually preferable to the use of average duration data which may be influenced by other forces. Second, when dealing with problems of labor market disequilibrium, equation (1) will not hold exactly, and one is not "required" to use D_u. Indeed, it seems more important to examine some hypotheses concerning influences on D_u describing the level of the statistic.

Barrett's discussion of this issue reflects at best a careless reading of the original article. For example, Barrett makes the remarkable statement: "To test the hypothesis that D_u does not vary between countries, Flanagan uses the rate of long duration unemployment, that is, the percent of the labor force unemployed fifteen weeks or more." No such hypothesis was offered, and the analysis of long duration unemployment rates was made for another purpose which is quite explicit in the original paper. Later, Barrett states ". . . his conclusion that international differences in duration are not significant . . ." (see p. 215) and still later "Flanagan contends that differences in the *U.S.* and Swedish unemployment rates are due entirely to differences in I_u" (p. 217). I drew no such conclusion in the paper. I did conclude that my analysis did not indicate relatively severe immobilities in the United States (which would have been reflected in part in a high average duration of unemployment) but did indicate ". . . a relatively high rate of separations and labor force entry along with the associated unemployment" (p. 130). This conclusion is in no way changed by Barrett's numbers.

A final distortion is found in the sentence, "It appears that Flanagan *found no difference* in the long duration unemployment rates (U_D) for the two countries because U_D contains an element of turnover which is higher for the United States" (p. 217, emphasis added). This, too, is an extremely imaginative statement in view of the data in the first full paragraph on page 129 of my article (as well as some of the numbers on page 128, which correspond to data provided in Barrett's Table 2).

Putting these difficulties aside, is it reasonable to conclude the D_u is relatively high in Sweden? Combining labor force survey data published in Statistiska Centralbyrån with assumptions about the mean duration of unemployment within broad duration categories which should yield estimates of D_u that are biased upward, I estimate the annual average duration of unemployment for both sexes to range between about 6.5 weeks and 8.5 weeks during the late 1960's. Although the year-to-year movements in these estimates correspond to those reported in Barrett's Table 1, the level is well below the estimates she reports for each sex. Of course, the most useful concept is that of completed duration. As noted earlier, however, for average completed duration to greatly exceed average duration as reported in the surveys, the unemployment continuation rate should decline as a spell of unemployment rate wears on. If this is the missing fact that reconciles my estimates with Barrett's, it is interesting to note that 1) it is the opposite of the behavior of unemployment continuation rates in the United States (see Kaitz), and 2) it does not accord well with Barrett's captive worker story.

The final issue regards the welfare implications of unemployment rate differentials. If Barrett feels that I arrived at strong conclusions regarding the welfare implications of *U.S.* or Swedish unemployment rates as a result of my analysis, she has seriously misread my article. Indeed, the final paragraphs of the article (p. 130) display a definite agnosticism concerning both the welfare implications of international unemployment rate differentials given the knowledge then available on the reasons for relatively high *U.S.* turnover and unemployment flows.

Barrett is considerably less cautious in a section that is distinguished by a relatively high ratio of speculation to evidence.[1] Most of Barrett's comments flow from the notion that the speed with which a worker finds a job may be an index of labor market efficiency, the observation of a relatively high D_u in Sweden, and an asserted captive worker effect. One important error in Barrett's reasoning is the presumption that spells of unemployment are always completed by accepting a job. The fact that some unemployment spells are terminated by labor force withdrawal weakens the use of D_u as an index of labor market efficiency. For example, within the United States, the *Current Population Survey* data report very similar average unemployment durations for black and white males. However, in an analysis of gross flow data, George Perry has found that the probability of terminating a spell of unemployment by labor force withdrawal is greater for blacks than for whites. Is the labor market for each race equally efficient?

Retaining the racial example, it is well known that behind the similar D_u is a black-white unemployment rate ratio of two to one. This implies a relatively large flow of new unemployment for blacks. Are the markets equally efficient? Recent analyses suggest not. These indicate that the practice of wage discrimination and occupational segregation tends to raise black turnover and associated unemployment, and that the probability of incurring unemployment, conditional on turnover, is higher for blacks. Readers may at least be able to agree with Barrett that a given unemployment rate is an imperfect measure of the welfare loss of joblessness.

REFERENCES

N. S. Barrett, "The U.S. Phillips Curve and International Unemployment Rate Differentials: Comment," *Amer. Econ. Rev.*, Mar. 1975, *65*, 213–21.

R. J. Flanagan, "The U.S. Phillips Curve and International Unemployment Rate Differentials," *Amer. Econ. Rev.*, Mar. 1973, *63*, 114–31.

H. B. Kaitz, "Analyzing the Length of Spells of Unemployment," *Mon. Lab. Rev.*, Nov. 1970, *93*, 11–20.

G. L. Perry, "Unemployment Flows in the U.S. Labor Market," *Brookings Papers*, Washington 1972, *2*, 245–78.

Statistiska Centralbyrån (*SCB*), "Arbetskraftsundersöningarna" (The Labor Force Surveys), *Statistiska Meddelanden*, various issues, 1968–71.

---

[1] In one instance, Barrett misreports regression results from Table 4 of my article in an effort to muster support for the "captive worker hypothesis." Contrary to Barrett's claim, the regressions reported in Table 4 show an inverse relation between the separation rate and the unemployment rate in the United States (as well as Sweden).

On The Reward System

By LIANG-SHING FAN*

The importance of the bonus structure in a public sector dominated economy has been noted by many writers. Since economic reforms in the *USSR* and many other socialist countries were initiated in the early 1960's, there have been a few studies analyzing the bonus schemes in Soviet-type economies. (See S. Gindin, Michael Ellman, Michael Keren, James A. Mirrlees, and Wayne A. Leeman.) The trend toward profit-related bonus schemes has been continuing in socialist economies, except China's industrial sector which completely dropped the bonus as a revisionist deviation.

There are two basic problems involved: one is the bonus which induces a greater effort than the normal exertion, and the other is the tautness (or slackness) of the plan that discourages (or encourages) the effort. Gindin noted the effort-inducing effect of the bonus and the fear of a potentially higher new plan target when the current plan is fulfilled or overfulfilled. His model explains the managers' and planners' conflicting interest, which is inherent in a planned economy.[1] Leeman discussed some bonus formulae in a Soviet-type economy. He noted that if the reward were given only to the fulfillment of physical output targets, managers would tend to ignore the volume, and if the reward were based on the volume, then managers would be indifferent to plans. However, he also noted that the bonus for overfulfillment induces the managers to understate their capacities. There is a conflict in the output-plan and profit-related bonus.[2] Ellman (1973) further analyzed the bonus based on the profit which reduces managers' incentive to understate the capacity. Ellman (1971) also made a detailed analysis of the institutional background of the *USSR* economy of the late 1960's that provides invaluable understanding of the Soviet reward system.

Keren's highly analytical paper emphasizes the tautness of plans and the effort-inducing bonus. Basically Keren tried to show that there exists, given a bonus, the highest permissible target, and also to show that there exists an "optimum size of bonus" which elicits the maximum exertion (effort) given the planned target.[3] Mirrlees, in a recent paper, demonstrated that when the uncertainty about agents' characteristics is ignored, a nearly optimal incentive system is "... to threaten dire punishment of the unlucky (or lazy) and give everyone else the same payment, the payment being just high enough to make sure that people are willing to offer themselves as agents despite the threatened penalties" (p. 5).

In this note, a bonus scheme which possesses some desirable characteristics is presented. It will be shown that managers operating under this reward system will have no incentive to understate or overstate their production capacity and they will strive to attain their target profits set by the planning authority.[4]

---

* Professor of economics, Colorado State University. This work was supported by a National Science Foundation grant to the University of Minnesota. I wish to thank Leonid Hurwicz for support and encouragement. I am also grateful to an anonymous referee who provided comments for the revision of this paper but I retain sole responsibility for its content.

[1] Gindin's utility function, pp. 147–48, is

$$u = f(B, T_n, T_0) \qquad \text{where } B = g(T, X)$$

$$\frac{\partial u}{\partial B} > 0, \qquad \frac{\partial u}{\partial T_n} < 0, \qquad \frac{\partial B}{\partial X} > 0$$

where B is the bonus. X the output, T_0 the current target, and T_n, the new target, are related to the current target as $T_n = T_0 + \mathrm{t}(X - T_0)$.

[2] See Leeman, pp. 444–45.

[3] Keren's managers have utility functions $u(x, c)$ where x is effort, c is consumption, and $u_c > 0$, $u_x \leqq 0$. However, the effort is induced by bonus as $\partial x/\partial b \geqq 0$ (see Keren, p. 477).

[4] In this analysis, there is always an uncertainty about the appropriateness of the output and input prices in the transition period for the calculation of profit. An implicit assumption is that the central plan-

I. The Bonus Structure

The exertion of effort depends upon the bonus structure $B(\pi)$ and the planned target profit π_0. It is the prospect of a potentially better reward which motivates managers and workers to work harder. The bonus can be higher when the planned target is fulfilled or overfulfilled; while there will be a danger of penalty when the target profit is not attained. Thus managers can insure a high bonus when the target is set low, but the low target and substantial overfulfillment during a current period usually mean a higher target in the next period. The profit can be higher if managers can minimize the cost of production. In this section, I try to analyze properties of a bonus scheme which seems to possess certain desirable characteristics.

The difficulty encountered by the planner in setting up the planned target for an economic unit is the tremendous informational requirements,[5] for example, necessary technological data of each unit. Depending upon the accuracy of the information provided by managers and the a priori information possessed by the planning office, the planned target may be "taut" or "slack." If we can devise a bonus scheme under which managers will have all the incentive to report the maximum (normal) production capacity and the incentive to fulfill the planned target, the informational requirements can be drastically reduced and the planned target will be (nearly) optimal or "reasonable" (neither taut nor slack).[6]

Let us define the bonus-penalty structure[7] as

$$(1) \qquad B(\pi) = \alpha\pi_0 + \beta(\pi - \pi_0) \qquad \alpha, \beta > 0$$

In the case where $\alpha = 0, \beta > 0$, the economic unit has the incentive to provide the planning authority false information about its normal production capacity. Thus π_0 will be set low in order that there will be a higher probability of overfulfillment of the plan. Subsequently, the bonus will be large. However, this incentive to cheat may be counterbalanced by the danger that π_0 for the (t+1)th period will be set much higher than that of the t-th period. In order to reduce the incentive to cheat on the reporting of the production capacity so that a high level of net benefit (profit) production is guaranteed, α is set high relative to β. In general $\alpha > \beta$, but it depends upon the underlying behavior with respect to risk aversion, etc.

Moreover, in order to avoid overstatement of the production capacity because of higher reward due to $\alpha > \beta$, we have to penalize the underfulfillment of the planned production. These measures are structured in such a way that the incentive to understate and the incentive to overstate the normal production capacity for the norm setting (π_0) of the planning authority is reduced.

In general we may write the reward system as[8]

$$(2) \qquad B = \alpha\{\pi_0 + [1 + \epsilon \operatorname{sign}(\pi_0 - \pi)] \cdot [\pi - \pi_0]\}$$

where $\epsilon > 0$ and ϵ will be preceded by the sign ($\pm$) depending upon $\pi_0 \gtrless \pi$. We may write this more specifically

(2a) In case of $\pi = \pi_0$, the bonus $B = \alpha\pi_0$ (target fulfillment)

(2b) In case of $\pi > \pi_0$, $B = \alpha[\pi_0 + (1 - \epsilon) \cdot (\pi - \pi_0)]$ (overfulfillment)

(2c) In case of $\pi < \pi_0$, $B = \alpha[\pi_0 + (1 + \epsilon) \cdot (\pi - \pi_0)]$ (underfulfillment)

The question to ask is whether (2) can eliminate the incentive to cheat on the information necessary in setting the target π_0 and whether the incentive to fulfill the target is preserved.

---

ning authority has approximated a proper set of socially desirable prices. This point was noted by E. G. Liberman, and others. I am grateful to the referee for clarifying this point.

[5] See Leonid Hurwicz (1969) for the informational efficiency of processes.

[6] Keren asserts that the optimum target is reached when given the bonus, the effort cannot be increased by raising the target, i.e., $\partial x/\partial Q^* = 0$ where Q^* is the target.

[7] It is similar to the one presented in Ellman (1973, p. 652).

[8] This formulation was suggested by Hurwicz.

A. *Deterministic Case*

If π^* is the maximum capacity of the ith unit and it is truthfully reported, then fulfillment of the plan will mean

$$(3)\qquad B = \alpha\pi^*$$

If the ith unit understates the production capacity by δ ($\delta > 0$), the result will be $\pi_0 = \pi^* - \delta$.

If the ith unit produces $\pi = \pi_0 = \pi^* - \delta$, the bonus will be

$$(4)\qquad B = \alpha[\pi^* - \delta] < \alpha\pi^*$$

If the ith unit produces at the maximum capacity π^*,

$$(5)\qquad \begin{aligned} B &= \alpha[\pi^* - \delta] + \alpha[(1-\epsilon)(\pi^* - \pi^* + \delta)] \\ &= \alpha\pi^* - \alpha\delta + \alpha\delta - (\alpha\ \delta\ \epsilon) < \alpha\pi^* \end{aligned}$$

If the ith unit produces less than the $\pi_0 = \pi^* - \delta$, say by γ,

$$(6)\qquad \begin{aligned} B &= \alpha[\pi^* - \delta] + \alpha[(1+\epsilon)(\pi^* - \delta \\ &\qquad - \gamma - \pi^* + \delta)] \\ &= \alpha\pi^* - \alpha\delta - \alpha\gamma - \alpha\epsilon\gamma \\ &= \alpha\pi^* - \alpha[\delta + (1+\epsilon)\gamma] < \alpha\pi^* \end{aligned}$$

Thus we may state that the bonus structure (2) will preserve the incentive to report and to fulfill the target (of the maximum capacity).

B. *Stochastic Case*

Suppose that π^* is the normal capacity of the ith unit and the probability is .5 that π^* is attainable, while there is a .25 probability that $\pi^* \pm \delta$ are attainable.

Then the expected value of the bonus when this norm is reported will be

$$(7)\qquad \begin{aligned} E(B) &= (1/2)\alpha\pi^* + (1/4)\alpha[\pi^* + (1-\epsilon) \\ &\qquad \cdot(\pi^* + \delta - \pi^*)] \\ &\qquad + (1/4)\alpha[\pi^* + (1+\epsilon) \\ &\qquad \cdot(\pi^* - \delta - \pi^*)] \\ &= \alpha\pi^* - (1/2)\alpha\epsilon\delta \end{aligned}$$

Case of Underreporting. If ith economic unit reported the condition so that the Planners set the norm π_0 as $\pi_0 = \pi^* - \mu > 0$, then the expected value of the bonus is

$$(8)\qquad \begin{aligned} \underline{E(B)} &= (1/2)\alpha[(\pi^* - \mu) + (1-\epsilon) \\ &\qquad \cdot\{\pi^* - (\pi^* - \mu)\}] \\ &\qquad + (1/4)\alpha[(\pi^* - \mu) + (1-\epsilon) \\ &\qquad \cdot\{\pi^* + \delta - (\pi^* - \mu)\}] \\ &\qquad + (1/4)\alpha[(\pi^* - \mu) + (1+\epsilon) \\ &\qquad \cdot\{\pi^* - \delta - (\pi^* - \mu)\}] \\ &= \alpha\pi^* - (1/2)\alpha\epsilon(\delta + \mu) \end{aligned}$$

Equations (7) and (8) will yield

$$E(B) - \underline{E(B)} = (1/2)\alpha\epsilon\mu > 0$$

Thus underreporting of the normal capacity will be bonus reducing. Thus there will be no incentive to underreport.

Case of Overreporting. Analogously, if the ith economic unit reports conditions (provides false information) so that the target is set to be $\pi_0 = \pi^* + \mu$, then the expected bonus will be

$$(9)\qquad \begin{aligned} \overline{E(B)} &= (1/2)\alpha[\pi^* + \mu + (1+\epsilon) \\ &\qquad \cdot(\pi^* - (\pi^* + \mu))] \\ &\qquad + (1/4)\alpha[\pi^* + \mu + (1-\epsilon) \\ &\qquad \cdot[\pi^* + \delta - (\pi^* + \mu)]] \\ &\qquad + (1/4)\alpha[\pi^* + \mu + (1+\epsilon) \\ &\qquad \cdot[\pi^* - \delta - (\pi^* + \mu)]] \\ &= \alpha\pi^* - (1/2)\alpha\epsilon(\mu + \delta) \end{aligned}$$

Equations (7) and (9) will yield

$$E(B) - \overline{E(B)} = (1/2)\alpha\epsilon\mu > 0$$

Thus there will be no incentive to overstate the normal production capacity. These simple illustrations indicate that the bonus structure (2) is "incentive" unbiased in the sense that individual units will have nothing to gain by under- or overreporting the normal production capacity.

II. Concluding Remarks

Ellman reported that the results of the establishment of such a reward system have not been totally satisfactory because of some uncertainties involved in the transitional period such as the appropriateness of the set of prices announced by the planning authority. Keren, p. 482, pointed out that the most important uncertainty is the delivery of inputs from other units. Liberman also reported the difficulty of the economic reform implementation because procedures of the measurement of profit are still evolving gradually. Nonetheless, the importance of the reward system has been firmly established in most of the socialistic economic systems, from Yugoslavia to the Soviet Union. However, the role of a well-devised reward system is not limited to planned economies. In any large-scale private enterprise where corporate planning is based on the information provided by numerous production units and in any public enterprise in market economies, the efficiency of production depends upon the reward system. The bonus structure (2) can be empirically tested in a given economic unit by varying the parameters α and ϵ so that a maximum effort-inducing formula can be approximated. A related fundamental issue of the existence of an incentive-compatible resource allocation mechanism was discussed in Hurwicz (1972, 1973a, b).

REFERENCES

M. Ellman, *Soviet Planning Today: Proposals for an Optimally Functioning Economic System,* Cambridge 1971.

———, "Bonus Formulae and Soviet Managerial Performance: A Further Comment," *Southern Econ. J.,* Apr. 1973, *39,* 652–53.

S. Gindin, "A Model of the Soviet Firm," *Econ. Planning,* 1970, *10,* 145–57.

L. Hurwicz, "On the Concept and Possibility of Informational Decentralization," *Amer. Econ. Rev. Proc.,* May 1969, *59,* 513–24.

———, "On Informationally Decentralized Systems," in C. B. McGuire and R. Radner, eds., *Decision and Organization,* Amsterdam 1972.

———, (1973a) "The Design of Mechanisms for Resource Allocation," *Amer. Econ. Rev. Proc.,* May 1973, *63,* 1–30.

———, (1973b) "On the Existence of Incentive-Compatible Allocation Systems," paper presented at the First Berlin Symposium on Planning, August 1973.

M. Keren, "On the Tautness of Plans," *Rev. Econ. Stud.,* Oct. 1972, *39,* 469–86.

W. A. Leeman, "Bonus Formulae and Soviet Managerial Performance," *Southern Econ. J.,* Apr. 1970, *36,* 434–45.

E. G. Liberman, *Economic Methods and the Effectiveness of Production,* New York 1973.

J. A. Mirrlees, "Optimal Incentive Schemes," paper presented at the First Berlin Symposium on Planning, August 1973.

The Industrial Composition of U.S. Exports and Subsidiary Sales to the Canadian Market: Comment

By DALE ORR*

In his paper in the March 1972 issue of this *Review*, Thomas Horst concludes ". . . tariffs imposed by the foreign country encourage *U.S.* firms to substitute subsidiary production for exporting" (p. 43). Initially this conclusion seems unquestionable. Horst did develop the micro-economic foundations which would lead to a negative relationship between tariffs and his measure of *U.S.* penetration. His estimation of this relationship across Canada's major group manufacturing industries provided strong statistical support.

The empirical relevance of tariffs to the foreign firm's production location decision is crucial to policy makers. After several studies based on questionnaires (see Herbert Marshall, Frank Southard, and Robert Taylor; A. E. Safarian; and D. Brash), Horst's paper is the first direct econometric investigation. This alone is justification for further study. Additionally, however, some of Horst's assumptions and procedures are questionable, and more detailed data has since become available. Reexamination of this issue is thus imperative.

I first consider alternative, and I allege more reasonable, treatments of excise taxes, sample size, measurement of tariff variables, and functional form of estimating equation. It will become apparent that these modifications only slightly weaken Horst's conclusion when the model is estimated at the major group level. However, the tariff is clearly insignificant when the same model is estimated at the more disaggregated three-digit level.

I shall use Horst's notation in what follows:

$X_i \equiv$ *U.S.* exports of commodity i to Canada, divided by total Canadian sales of i

$P_i \equiv$ Sales of commodity i in Canada by *U.S.*-owned subsidiaries, divided by total Canadian sales of i

$S_i \equiv X_i + P_i$

$Z_i \equiv$ ratio of Canadian market size to *U.S.* domestic production of i

$t_i \equiv$ nominal tariff rate on imports of i

$e_i \equiv$ effective tariff rate on imports of i

I. A Reexamination of Horst's Analysis: Estimates at the Major Group Level

I first question Horst's inclusion of excise taxes within his measure of the tariff rate. These excise taxes were (unknown to Horst at the time) inappropriately included in James Melvin and Bruce Wilkinson's original measures of the tariff rate. The foreign firm would pay the tax regardless of whether it supplied the Canadian market via a Canadian subsidiary or imports by Canada. Therefore excise taxes should not be expected to explain cross-section differences in X/S. Equation 1) of Table 1 repeats Horst's basic results (p. 41) where excise taxes were included in the nominal $(1+t_1)$ and effective $(1+e_1)$ tariff rates for the beverage and tobacco industries. When these excise taxes are subtracted out, neither tariff variable is as negatively associated to the import-total penetration ratio (X/S), (see equation 2)).

Next I examine the results of dropping and combining industries for Horst's sample of seventeen major group industries, rather than the complete set of twenty. I take data on Canada's imports from the United States, production by *U.S.* controlled corporations in Canada, and tariff rates for each of the seventy-one well-defined three-digit manufacturing industries.[1] These three-digit in-

---

* Department of economics, University of British Columbia. I wish to acknowledge the helpful comments of my colleagues Russ Uhler and Alan Woodland. I am also grateful to Thomas Horst who kindly provided his data and details on several statistical procedures.

[1] Sources of data at the three-digit level are: 1) Imports by Canada from the United States (X): Com-

TABLE 1—THE IMPACT OF TARIFFS ON THE METHOD OF *U.S.* PENETRATION: ESTIMATES OVER THE MAJOR GROUP INDUSTRIES

Equations	Imports/total penetration variable	Constant	Tariff Variable	Canada/*U.S.* cost of production advantage	R^2	N
1)	$\ln (X/S)_H$	$-.23$	$-2.45 \ln (1+t_1)^c$	$-12.6\ Z^c$	.696	17
	$\ln (X/S)_H$	$-.38$	$-1.20 \ln (1+e_1)^c$	$-11.3\ Z^c$	.711	17
2)	$\ln (X/S)_{HI}$	.32	$-5.28 \ln (1+t_2)^b$	$-16.1\ Z^c$	.609	17
	$\ln (X/S)_{HI}$	.39	$-3.77 \ln (1+e_2)^b$	$-15.7\ Z^c$	.636	17
3)	$\ln (X/S)$	.63	$-4.95 \ln (1+t_3)^b$	$-23.5\ Z^c$	.722	17
	$\ln (X/S)$	.68	$-3.75 \ln (1+e_3)^c$	$-22.3\ Z^c$	.753	17
4)	$\ln (X/S)$	.37	$-5.32 \ln (1+t_4)^a$	$-21.02\ Z^c$	.548	20
	$\ln (X/S)$	.27	$-3.28 \ln (1+e_4)^b$	$-19.85\ Z^c$	.541	20
5)	X/S	1.29	$-\ .77\ \ (1+t_4)^b$	$-\ 2.59\ Z^c$	.428	20
	X/S	.96	$-\ .45\ \ (1+e_4)^b$	$-\ 2.43\ Z^c$	.427	20

[a] Significant at the .90 level, one-tail test.
[b] Significant at the .95 level, one-tail test.
[c] Significant at the .99 level, one-tail test.

dustries are then combined into their appropriate major group industries. Those major group industries that are dropped or combined by Horst are then treated likewise to give seventeen observations quite comparable to Horst's. Horst's value of Z will be used in all regressions. The prime difference between my data set and Horst's is that my data on major group industries are based only on their well-defined three-digit components. Horst's data also include the manufacturing activity in the three-digit industries identified as "other" or "miscellaneous." These usually account for only about 15 percent of total Canadian manufacturing. It is not surprising then that our variables are highly correlated to their counterpart in Horst's data. Horst's and my measures of X/S $(1+t, 1+e)$ are correlated .85 (.98, .98). Comparison of equations 2) and 3) indicates that my data set yields results quite consistent with Horst's.

Horst's equations are estimated over a sample of seventeen, rather than the complete set of twenty major group industries. He dropped the food industry and combined the textile, clothing and knitting mills industries, partly because of difficulties with the import data. As a measure of the major group tariff he weighted the tariffs of the three-digit components by *total* imports by Canada. Since the dependent variable is based on *U.S.* imports, the tariff rates should more appropriately be weighted by imports only from the United States (if, indeed, imports is the relevant weighting factor). Weighting the three-digit tariffs by the data on *U.S.* imports not only permits construction of more appropriate tariff data at the major group level, but permits expansion of the sample to the complete set of twenty major group industries. Comparison of equation 4) with equation 3) reveals the tariff variables are still negative and significant after this change in calculation of tariff rates and sample size.

---

modity Trade Statistics, U.K. 1967. Classifications, calculations and conversion from commodity to three-digit data kindly provided by Gary Hewitt. 2) Production by *U.S.* controlled corporations in Canada (P): $X+P=S$. Unpublished data provided by CALURA. 3) Tariff rates t, e, Melvin and Wilkinson. The sample of seventy-one includes all those three-digit industries except those labelled "other" or "miscellaneous." It should be stressed that my access to unpublished data on *U.S.* control at the three-digit level was crucial to this study. These data were unavailable to anyone at the time Horst was preparing his paper.

These equations take *logs* of variables which are already percentages (i.e., X/S, $1+t$, $1+e$). Initially one might expect that the relationship between X/S and the tariff would be linear in percentage point differences. This specification would, of course, be tested by regressing X/S on $1+t$, or $1+e$. It seems this specification also appealed to Horst a priori. He writes that this "... logarithmic transformation was not specified a priori but only after the tendency of export [Canada's imports] shares to respond much more sharply to an increase in a low tariff than to an increase in a high tariff was observed" (p. 41). He did use the linear, rather than the *log* form, in examining this same problem for the United Kingdom and Common Market (see p. 43).

From the observation that the R^2 is higher in the *log* than the linear form, it is not accurate to conclude the *log* specification is more appropriate. These equations are predicting different quantities, i.e., $\ln X/S$ and X/S. To compare the performance of the *log* to the linear specification we calculate the values of $\ln X/S$ predicted by equation 4), i.e., $\ln X/S$. Then

$$\widehat{e^{\ln X/S}}=X/S$$

Correlating X/S and X/S, then squaring this total yields the R^2 of X/S implied by the *log* relationship of equation 4). This R^2 implied by the *log* relationship of equation 4) is greater than that of the linear relationship (equation 5)) for both nominal (.658>.428) and effective (.653>.427) tariff rates. This supports Horst's contention that the *log* specification, although it has less theoretical appeal, fits the set of data better than the linear specification.

Since the sample size is so small these results could be heavily influenced by several outlying observations. We note from the scatter diagram of each equation in Table 1 that the printing and publishing and machinery industries are industries with particularly high values of the dependent variable and low tariffs. These two outlying observations, therefore, can be expected to exert a strong influence towards negative estimates of the tariff coefficients. When these two observations are dropped from equation 2) based on Horst's data, the t-statistics on the tariff variables are barely significant at the .90 level.[2] Dropping these two observations from any of the equations based on our data (equations 3), 4), and 5)) leaves the tariff variables negative but insignificant.

From Table 1 we conclude that insertion of alternative assumptions only slightly weakens the significance of the tariff variable when the model is estimated over the major group industries. We proceed next to disaggregate the data and estimate this model at the three-digit level.

II. Estimates of the Model at the Three Digit Level

Disaggregating these twenty major group industries into their three-digit components provides the results of equations 1) and 2), Table 2.[3] Compare these results to any of the equations in Table 1, estimated over the major group industries. We note the tariff variables are not even negative, let alone significantly negative, determinants of the import-total penetration ratio when the same relationship is estimated at the three-digit, rather than major group level.

The negative significance of tariffs at the major group, but not three-digit level, is not due to arbitrary aggregation procedures, measurement error, or multicollinearity. The comparison of equations 3) and 4) in Table 1 indicated the tariff variables were significantly negative regardless of what import data were used as weights in the aggregation procedure. The major group data on X and S were constructed simply by summing the appropriate three-digit components. For some of the three-digit industries calculation of Z is quite difficult since there is no direct *U.S.* counterpart to the Canadian industry. However, Consumer and Corporate Affairs list direct *U.S.* counterparts for forty-seven

---

[2] The t-statistic on the nominal (effective) tariff rate falls from 2.24 to 1.45 (2.53 to 1.49). One-tail $t_{.90}$ with 12 degrees of freedom = 1.36.

[3] Three industries had to be dropped from the sample to estimate the *log* specification (equation 1)) since $X/S=0$.

TABLE 2—THE IMPACT OF TARIFFS ON *U.S.* PENETRATION: ESTIMATES OVER THE THREE-DIGIT INDUSTRIES

Equations	Imports/total penetration variable	Constant	Tariff Variable	Canada/*U.S.* cost of production variable	R^2	N
1)	$ln\,(X/S)$	-2.12	$+2.05\,ln\,(1+t_5)$	$-.42\,Z$	.014	68
	$ln\,(X/S)$	-2.14	$+1.40\,ln\,(1+e_5)$	$-.65\,Z$	.029	68
2)	X/S	.046	$+\ .21\,(1+t_5)$	$-.36\,Z$	.010	71
	X/S	.216	$+\ .06\,(1+e_5)$	$-.40\,Z$	.008	71
3)	X	$.30(S)$	$-\ .15\,(1+t_5)\mathrm{x}S$	$+.93\,(Z\,\mathrm{x}\,S)^{a}$	.402	71
	X	$.19(S)$	$-\ .05\,(1+e_5)\mathrm{x}S$	$+.94\,(Z\,\mathrm{x}\,S)^{a}$	.400	71

[a] See fn. b, Table 1.

of these seventy-one industries. Estimation of equation 2), Table 2 over only these industries also yielded positive but insignificant signs on the tariff variables. The difference in results between major group and three-digit levels is not due to multicollinearity between Z and the tariff variables. At each level of aggregation the correlation between Z and the tariff variables is negligible. Over the major group industries $1+t_4$, $(1+e_4)$ is highly negatively correlated to X/S, $r=-.39$ ($r=-.44$). At the three-digit level these correlations are slightly positive, $r=.07$ and .05, respectively.

Since the linear form (Table 2, equation 2)) fits the three-digit data set virtually as well as the *log* form (equation 1)), but is preferred on conceptual grounds, I will use equation 2) as the basis for further analysis. Examination of the residuals from equation 2) reveals that the largest residuals are predominantly positive residuals and they occur in industries with highest values of X/S. These values in turn can usually be traced to unusually low values of S. This heteroscedasticity, coupled with considerations of measurement error and theoretical reasons, suggests a weighted regression rather than equation 2) may be a more appropriate test of the hypothesis.[4] However, in the implied weighted regression (equation 3)) the tariff rates were still statistically insignificant. I conclude that higher tariffs discourage imports but there is no evidence that *U.S.* controlled production in Canada is substituted for these imports.

III. Conclusion

This paper reexamines the analysis and conclusions of Horst's pioneering study on the impact of tariffs on the method of *U.S.* penetration. Horst concluded that tariffs are a highly significant factor in explaining whether *U.S.* firms choose to penetrate the Canadian market by Canadian imports rather than *U.S.* controlled production in Canada. His conclusion is based on estimates across the Canadian major group manufacturing industries. The principle finding of my paper is that whether tariffs are an empirically relevant determinant of the method of *U.S.* penetration depends crucially on the level of aggregation of industries chosen for analysis. Insertion of several assumptions and procedures alternative to Horst's left his conclusion only slightly weakened as long as his model is estimated at the major group level. However, when this same model is estimated at the more disaggregated three-digit level the tariff rate is clearly insignificant.

---

[4] Since the variance of the residuals is larger for industries with small values of S this suggests heteroscedasticity is present. Measurement error, especially rounding error, would affect measured X/S relatively more seriously for small than large values of S. Also, the researcher may want industries with low values of S to carry less weight in the estimated relationship. That is, if $X=.10$, $P=.40$, $X/(X+P)=X/S=.20$; that may represent a more important and more explicit choice of of the method of *U.S.* penetration than the industry where $X=.01$, $P=.04$, $X/S=.20$. In equation 2), however, each of these industries carries equal weight in determining the overall relationship. It appears that the variance of the error term increases as $1/S^2$ (rather than $1/S$ or $(1/\sqrt{S})$). In this case the appropriate weight is S. Weighting by S, of course, deemphasizes the role of those industries with small values of S. These are the industries with the largest residuals, most subject to measurement error, and "least important." The fact that the industries with largest residuals have particularly low S indicates that some variable which could explain low *U.S.* penetration has been omitted from equation 2). We expect *U.S.* penetration to be particularly low in industries where the relevant market is local or regional. However, addition of a dummy variable for regional industries still left the tariff variables insignificantly different from 0. It is interesting to note however that when X (import intensity) was regressed on these variables each had the expected sign. This may instill some confidence in the data and estimates at the three-digit level.

Econometricians would not be surprised that the results are sensitive to the level of aggregation. For example, see Henri Theil (ch. 11) or Yehuda Grunfeld and Zvi Griliches. Reasons offered for sometimes preferring estimates at the more aggregate level are difficulty in specifying micro behavior, more measurement error at the disaggregated level, and desire to implement policy at the aggregate level. These reasons do not seem to support aggregation in this particular case. The major group industries contain such a diversity of products it seems more difficult to postulate a behavioral relationship between the average level of variable at this level. Tariff and import data are reported at the commodity level; foreign control at the three-digit level. Thus there appears no particular reason to suspect more measurement error at the three-digit than major group level. Recall that even when those industries likely most subject to measurement error were discounted, the tariff was still insignificant. Policy decisions regarding tariffs are most often taken at the commodity level. The decision regarding the optimum method of foreign penetration is, of course, taken at the firm level. For this particular issue, it therefore seems reasonable to place more confidence in the results at the more disaggregated level.

My results should not be interpreted as refuting Horst's conclusion, but rather calling it to question. The Horst paper is a pathbreaking study on a very relevant issue. Hopefully, my paper, like Horst's, will stimulate others to further investigation.

REFERENCES

D. Brash, *American Investment in Australian Industry*, Cambridge, Mass. 1966.

Y. Grunfeld and Z. Griliches, "Is Aggregation Necessarily Bad?" *Rev. Econ. Statist.*, Feb. 1960, *42*, 1–13.

T. Horst, "The Industrial Composition of U.S. Exports and Subsidiary Sales to the Canadian Market," *Amer. Econ. Rev.*, Mar. 1972, *62*, 37–45.

H. Marshall, F. Southard, and R. Taylor, *Canadian-American Industry: A Study in International Investment*, New Haven 1936.

J. Melvin and B. Wilkinson, "Effective Protection in the Canadian Economy," Special Study No. 9, Economic Council of Canada, Ottawa 1968.

A. Safarian, *Foreign Ownership of Canadian Industry*, Toronto 1966.

H. Theil, *Principles of Econometrics*, New York 1971.

Consumer and Corporate Affairs, Government of Canada, *Concentration in the Manufacturing Industries of Canada*, Ottawa 1971.

The Industrial Composition of U.S. Exports and Subsidiary Sales to the Canadian Market: Reply

By THOMAS HORST*

Dale Orr's comment offers several valuable and interesting criticisms of my paper. Rather than magnifying differences of judgment and interpretation of our findings, let me focus on the apparent failure of his three-digit industry data to confirm my findings based on two-digit industry statistics. This failure sheds some interesting light on the pitfalls of cross-industry analysis and indirectly on the broader issue of how tariffs affect foreign trade and investment patterns.

However simple and straightforward the analysis of tariffs may be at a theoretical level, trying to show their impact empirically is no easy task. Trade and investment patterns reflect much more than tariffs. The size and rate of growth of the recipient market, the transport and inventory cost of exporting, the availability of capital, labor, raw materials and support services, the size of start-up and overhead costs of a foreign subsidiary, the requirements for marketing the product locally and the competition from other producers, domestic and foreign, all impinge upon trade and investment flows. Firms' behavior is often constrained by past commitments, so that current behavior reflects not only current conditions, but past histories as well. Finally, the willingness of tariff setters to give more protection to local producers threatened by import competition can lead to a perverse statistical relationship between observed tariff and import levels. Although econometric solutions to the problems of simultaneous determination, lagged adjustment and these other problems, exist, the available data are too skimpy and flawed to permit the effective use of sophisticated methods.

Normally, when one moves from a higher to a lower level of aggregation, the statistical significance of estimated coefficients improves—if nothing else, one has more degrees of freedom. Accordingly, Orr's finding to the contrary is rather surprising, but not completely implausible. While there are several possible explanations for this result, in this particular instance the deterioration in statistical significance appears to be due partly to errors of misclassification and mismatching at the three-digit level, but mainly to the implicit difference in weights given to various industries.

Let me be more specific. The negative and statistically significant coefficient of the tariff variable using the more aggregate data is due to a few industries (machinery and publishing) where tariffs are lowest and *U.S.* exports largest and to a few industries (beverages, tobacco, and rubber) where tariffs are highest and *U.S.* exports smallest. Over the intermediate range, the relationship between tariffs and *U.S.* exports appears haphazard.

In disaggregating from the two- to the three-digit level, the five extreme industries were decomposed into only nine industries, while the twelve intermediate industries accounted for fifty-two industries in the new scheme. The disaggregation was implicitly a reweighting of the sample points, and an unfavorable one at that. While one could quibble with Orr on which level of aggregation is to be preferred a priori, the real question is not really one of choosing the optimal level of aggregation. Rather, the point is that in a crude, cross-industry analysis, the impact of tariff policy can be seen only in the extreme cases. While tariffs may make a difference over the intermediate range, their influence is difficult to disentangle from other considerations.

What then should we conclude about the role of tariffs in influencing *U.S.* exports and subsidiary sales to Canada and elsewhere? As Orr acknowledges, support for the notion that tariffs can effect a substitution of subsidiary production for exports rests on a far broader basis than our econometric investigations. But both published research and common sense indicate that the actual relationships are complex, and to the extent that my original paper oversimplified the issue, Orr's comment has set the record straight.

---

* The Brookings Institution.

Black Education, Earnings, and Interregional Migration: A Comment and Some New Evidence

By CHARLES R. LINK*

In a recent article in this *Review*, Leonard Weiss and Jeffrey Williamson (hereafter referred to as W-W) examined the impact of education and several other factors on the earnings of a sample of black males. Utilizing multiple regression analysis and data from the 1967 Survey of Economic Opportunity (*SEO*) they find that . . . "Education has a strong and consistent effect on black incomes for the sample as a whole, and for each age group in the interaction equation" (p. 379). W-W's results differ from earlier studies using 1960 Census data that find practically no effect of black education on black incomes (see Giora Hanoch, Randall Weiss, Lester Thurow).

In order to test for the effects of full employment and, more importantly, parameter shifts that may have occurred during the 1960's, W-W apply Thurow's human capital function to their 1967 *SEO* data. They compare their estimates with Thurow's and conclude: "Our study also seems to show a significant shift in the earnings function for blacks during much of the past decade" (p. 382). They find that black incomes in 1967 appear to be at least as responsive to education as white incomes in 1960 with the increases appearing at all education levels. Finally, large increases in the impact of experience on earnings are found for young blacks (pp. 382–83).

This brief note presents further work which sheds light on whether the results obtained by W-W represent a long-term trend. Thurow's human capital function is reestimated utilizing 1970 Census data and the results compared with W-W and Thurow.

Thurow's human capital function takes the form

$$(1) \qquad Y = A\, Ed^{b} Ex^{c}$$

where an individual's income (Y) is dependent on his number of years of education (Ed) and experience (Ex), the shift coefficient (A), and the elasticities of income with respect to education b and experience c.[1] The shift coefficient picks up the effect of other relevant factors such as discrimination, market imperfections, and unionization, pp. 73–74.

To test the hypothesis that income elasticities vary for different ranges of education, Thurow disaggregates the continuous education variable into three continuous variables for the classes 0–8, 9–12, and more than 12 years of education. Experience is analogously divided into four continuous variables representing 0–5, 6–15, 16–35, and more than 35 years of experience. The disaggregated function is

$$(2) \qquad Y = A\, Ed_1^{b_1} \ldots Ed_3^{b_3} Ex_1^{c_1} \ldots Ex_4^{c_4}$$

Following Thurow, median income, age, and education data are taken directly from the 1970 Census for males aged eighteen to sixty-four and used to reestimate the human capital function. Experience is calculated by subtracting one's age of entry into the labor force from his present age. It is assumed that males with a high school education or less enter the labor force at age eighteen and college graduates at age twenty-two (see Thurow, pp. 75–76). Functions are estimated for six different groups: all black; all white; North white; North black; South black; and

---

* Assistant professor of economics, University of Delaware. This research was made possible by a General Faculty Research Grant at the University of Delaware. I would like to acknowledge the helpful comments of Ed Ratledge, George Borts, and a referee on an earlier draft of the paper. Helpful comments were also given by several of my colleagues at the University of Delaware. Of course, any errors are my sole responsibility.

[1] See Thurow, pp. 189–90, for a critique of the human capital function.

South white. The Census data provide eighty-six observations for each of the six groups.[2]

In the regression analysis the human capital function for each group is estimated in the form:

$$(3)\quad \ln Y = A + b_1 \ln Ed_1 + b_2 \ln Ed_2 + b_3 \ln Ed_3 + c_1 \ln Ex_1 + c_2 \ln Ex_2 + c_3 \ln Ex_3 + c_4 \ln Ex_4$$

where $Y = 1970$ median income, the Ed_i $(i = 1, 2, 3)$ are continuous variables indicating years of education to a maximum of 8, 12, and more than 12 years, respectively. Similarly, Ex_k $(k = 1, 2, 3, 4)$ represents 0–5, 6–15, 16–25, and more than 35 years of experience. Thurow omitted terminal education and experience variables whenever the standard errors exceeded their regression coefficients (pp. 72–76 and 187–88). In four of the six equations estimated in the present study, c_4 was less than its standard error and in the other two equations it was only slightly larger than its standard error. To enhance comparability with Thurow, Ex_4 is thus excluded from the 1970 regression equations reported in Table 1.[3]

Income elasticities for various levels of education and experience are calculated by adding b or c coefficients. The elasticities of income with respect to 0–8, 9–12, and more than 12 years of education are $b_1+b_2+b_3$, b_2+b_3, and b_3, respectively. Similarly elasticities of income with respect to 0–5, 6–15, 16–35, and more than 35 years of experience are $c_1+c_2+c_3+c_4$, $c_2+c_3+c_4$, c_3+c_4, and c_4, respectively (see Thurow, p. 187).[4] The estimated human capital functions for 1960, 1967, and 1970 are presented in Table 1 and the resulting income elasticities are shown in Table 2.[5]

Income elasticities of education for a particular education range should be interpreted as the percentage change in income arising from a one percentage point change in education. For instance, the income elasticity of education for all whites who attained 0–8 years of education in 1960 is .11. In this case a one percentage point increase in education is associated with an increase in income of .11 percent. Similar interpretations are given to the income elasticities for experience. Income elasticities are calculated for blacks and whites in the total United States, the North and the South.

W-W state that "Elasticities of income with respect to education [for blacks] have increased in *all* regions and *all* education classes" (p. 381). However, examination of the results for 1970 indicates that in absolute

---

[2] Thurow's analysis included 72 age-education categories. The increased number of observations in the 1970 data is due to the finer classifications of educational attainment. Each of the eight age groups with the exception of the youngest, has 11 educational categories (e.g., no school completed, 1–3 years, etc.) each of which has a corresponding median income figure. The youngest age group has only 9 educational categories. Thus the total number of observations is 86. In a small number of age-education categories for blacks there were not enough subjects to calculate a median income figure. Thus the number of observations is less than 86 and degrees of freedom are slightly lower in the black vs. white regressions.

[3] Exclusion of the more than 35 years of experience variable causes only minor changes in the education coefficients and virtually no changes in the income elasticities computed below.

[4] Examples of the way in which the data are coded for the regression analysis will give the reader an insight into why the elasticities are calculated as indicated above. Recall that Ed_1, Ed_2, and Ed_3 are continuous variables that indicate the number of years of education to maxima of 8, 12, and more than 12, respectively. Consider the cases of individuals with 3, 11, and 16 years of education. In the first case the coding is $Ed_1 = 3$, $Ed_2 = 3$, and $Ed_3 = 3$. The individual with 11 years is coded as $Ed_1 = 8$, $Ed_2 = 11$, and $Ed_3 = 11$. Finally our college graduate is coded as follows: $Ed_1 = 8$, $Ed_2 = 12$, and $Ed_3 = 16$. Coding the data in this fashion allows us to estimate the income elasticity associated with any particular education range as the sum of the coefficient for those who stopped at that level and the coefficient(s) associated with individuals who went on to attain higher levels of education. Inclusion of the latter coefficients allows for the effect of the first n years on the incomes of individuals who surpassed this level of education. Analogous reasoning is utilized in the coding of the Ex variables.

[5] Resulting $\bar{R}^2$s for 1960 and 1970 are higher as a result of using aggregate data. W-W use individual data. A Chow test was used to check for structural differences in the North and South human capital functions for both blacks and whites in 1970. The hypothesis of the same structure was rejected at the 1 percent level.

Table 1—Regression Results for Thurow's Human Capital Function (for Males)[a]

		Education			Experience				
	Constant-A	b_1	b_2	b_3	c_1	c_2	c_3	c_4	$\bar{R}^2$
				Thurow Results[b] (1960)					
Nonwhite									
All	6.6633	−0.6839	−0.5664	1.3348	−0.3897	0.7881	−0.1378	[c]	.89
	(0.0573)	(0.1919)	(0.3907)	(0.2480)	(0.0949)	(0.1638)	(0.1015)		
North	6.9208	−0.4533	−0.7074	1.2235	−0.3557	0.7365	−0.1089	[c]	.89
	(0.0515)	(0.1725)	(0.3512)	(0.2230)	(0.0853)	(0.1473)	(0.0913)		
South	6.5635	−0.4908	−1.1655	1.7374	−0.3693	0.8712	−0.2598	[c]	.89
	(0.0528)	(0.1768)	(0.3599)	(0.2285)	(0.0875)	(0.1509)	(0.0935)		
				Weiss-Williamson Results (1967)					
Black									
All	3.7421	−1.5902	0.0236	1.7718	0.9073	0.6942	4.0486	−4.0490	.143
	(0.1692)	(0.2221)	(0.5609)	(0.4496)	(0.2186)	(0.2312)	(0.4137)	(0.3356)	
North	3.8266	−1.2579	−0.0453	1.5086	1.1588	0.1777	5.5192	−5.1526	.141
	(0.2693)	(0.3491)	(0.7998)	(0.6309)	(0.3289)	(0.3507)	(0.6935)	(0.5804)	
South	3.8598	−1.4728	−0.5627	2.2222	0.7295	1.0891	3.1310	−3.4439	.136
	(0.2262)	(0.3031)	(0.8045)	(0.6444)	(2.9225)	(0.3070)	(0.5177)	(0.4106)	
				Link Results (1970)					
Black									
All	7.4525	−0.7371	−0.9894	1.7747	−0.4594	0.8624	−0.2458	[c]	.901
	(0.3491)	(0.1816)	(0.3531)	(0.2130)	(0.0789)	(0.1492)	(0.0931)		
North	7.6931	−0.4331	−1.2786	1.7395	−0.5055	0.9251	−0.2531	[c]	.887
	(0.3639)	(0.1893)	(0.3680)	(0.2220)	(0.0823)	(0.1555)	(0.0970)		
South	7.3639	−0.5631	−1.3136	1.9324	−0.3880	0.8098	−0.2579	[c]	.897
	(0.3463)	(0.1806)	(0.3537)	(0.2142)	(0.0795)	(0.1478)	(0.0921)		
White									
All	7.6556	−0.3513	−1.7385	2.1568	−0.6150	0.9461	−0.1785	[c]	.882
	(0.4311)	(0.2213)	(0.4272)	(0.2568)	(0.0954)	(0.1823)	(0.1139)		
North	7.7270	−0.0274	−2.2387	2.3240	−0.7031	1.0200	−0.1755	[c]	.853
	(0.4900)	(0.2516)	(0.4856)	(0.2919)	(0.1084)	(0.2073)	(0.1294)		
South	7.5363	−0.5337	−1.5459	2.1650	−0.5272	0.8550	−0.1654	[c]	.887
	(0.4318)	(0.2217)	(0.4279)	(0.2572)	(0.0955)	(0.1826)	(0.1141)		

[a] The 1960 results are from Thurow, Table G-1, p. 188. The 1967 results are from Weiss and Williamson's Table 6, p. 380. The estimation equation is given in the text. Figures in parentheses are standard errors.
[b] The nonwhite category in 1960 included certain other races in addition to blacks.
[c] Not estimated for 1960 or 1970 (see text).

terms the black gains are not as impressive as W-W claim.[6] In seven of the nine potential comparisons, 1970 elasticities are less than those for 1967.

In relative terms, although the 1970 income elasticities for blacks in the 13 or more years of education range have indeed increased from their 1960 values in all regions of the country and are similar to those found by W-W in 1967, substantial increases have also occurred in income elasticities for whites with a college education. Consequently, blacks with 13 or more years of education have experienced only small gains relative to whites. The ratio of college white to college black elasticities in 1960 and 1970 is 1.30 and 1.22, respectively.

---

[6] The portion of the increase in black returns to education and experience between 1960 (i.e., Thurow) and 1967 (i.e., W-W) which is real and that which is statistical artifact is unclear. In other words, given the well-known fact of a skewed distribution of income, the use of individual data fitted through mean incomes to estimate the human capital function may or may not yield substantially different results than similar data which is aggregated and fit on median incomes.

TABLE 2—INCOME ELASTICITIES OF EDUCATION AND EXPERIENCE FOR MALES, BY COLOR AND REGION, 1960, 1967, AND 1970[a]

		Years of Education			Years of Experience				Shift Coefficient[c]
Sample		0–8	9–12	13+	0–5	6–15	16–35	35+[b]	
All white	(1960)	0.11	0.72	1.73	0.20	0.71	−0.09		$1140
	(1970)	0.07	0.42	2.16	0.15	0.77	−0.18		2112
All nonwhite	(1960)	0.08	0.76	1.33	0.26	0.65	−0.14		783
All black	(1967)	0.21	1.80	1.77	1.60	0.69	0.00	−4.05	—
	(1970)	0.05	0.79	1.77	0.16	0.62	−0.25		1724
Northern white	(1960)	0.10	0.52	1.70	0.20	0.70	−0.06		1278
	(1970)	0.06	0.09	2.32	0.14	0.84	−0.18		2269
Northern nonwhite	(1960)	0.06	0.51	1.22	0.27	0.63	−0.11		1013
Northern black	(1967)	0.21	1.46	1.51	1.70	0.54	0.37	−5.15	—
	(1970)	0.03	0.46	1.74	0.17	0.67	−0.25		2194
Southern white	(1960)	0.11	0.90	1.91	0.20	0.75	−0.21		702
	(1970)	0.09	0.62	2.16	0.16	0.69	−0.17		1875
Southern nonwhite	(1960)	0.08	0.57	1.74	0.24	0.61	−0.26		709
Southern black	(1967)	0.19	1.66	2.22	1.51	0.78	−0.31	−3.44	—
	(1970)	0.06	0.62	1.93	0.16	0.55	−0.25		1578

[a] The 1960 figures are from Thurow, Table 5-1, p. 77. The 1967 elasticities are from Weiss and Williamson's Table 7, p. 381. The 1970 elasticities are calculated from Table 1.

[b] Not calculated for 1960 or 1970 (see text).

[c] Antilog of constant term (*A*).

The big relative gains have accrued to blacks in the 9–12 years of education category. Even here the absolute gains for blacks were modest between 1960 and 1970 but the corresponding white income elasticities dropped drastically. The South excepted, blacks in 1970 had larger elasticities than whites in the 9–12 years of education range.

With respect to primary education, all of the elasticities for blacks and whites alike fell during the decade of the 1960's. This deterioration reflects the competitive disadvantage confronting such males as a result of limited marketable skills and possibly a lack of credentials.

The 1970 data do not show the large increase in the responsiveness of income to experience for young blacks found by W-W, p. 381. The elasticities of income with respect to experience indicate, if anything, that whites have gained relative to blacks in the 1960's.

It is undoubtedly true that the tight labor market during the 1960's is partly responsible for the modest but positive relative gains received by blacks during the 1960's. W-W, p. 381, also give credit to legislation designed to eliminate discrimination. As was mentioned earlier the shift coefficient (*A*) reflects, among other factors, the impact of discrimination on earnings. Shift coefficients, derived by taking the antilog of the constant term (*A*), for the 1970 and 1960 models are shown in Table 2. In the North and for the whole United States, the shift coefficients for blacks although still smaller than the ones corresponding to whites have increased relative to those for whites. This is consistent with the hypothesis that civil rights legislation has at least to some degree enhanced the relative economic position of blacks.

In conclusion, even though investments in education and training have undoubtedly played a significant role in helping to improve the position of blacks during the past decade, the gains are not as large as earlier data indicated. It appears that policies in addition to human capital investments will be necessary if we are to bring about larger improvements in the economic position of blacks. Examples of such policies include a negative income tax, job creation programs, and the reduction of imperfections in the labor market brought about by factors such as racial discrimination.

REFERENCES

G. Hanoch, "Personal Earnings and Investment in Schooling," unpublished doctoral dissertation, Univ. Chicago 1965.

L. Thurow, *Poverty and Discrimination,* Washington 1969.

L. Weiss and J. G. Williamson, "Black Education, Earnings, and Interregional Migration: Some New Evidence," *Amer. Econ. Rev.,* June 1972, *62,* 372–83.

R. D. Weiss, "The Effect of Education on the Earnings of Blacks and Whites," *Rev. Econ. Statist.,* May 1970, *52,* 150–59.

Office of Economic Opportunity, "Survey of Economic Opportunity," (*SEO*) conducted spring 1967, available on tape, Data Bank, Univ. Wisconsin.

U.S. Bureau of the Census, *U.S. Census of Population:* 1970, Subject Reports, Educational Attainment, Final Report PC(2)-5B, Washington 1973.

Black Education, Earnings and Interregional Migration: Even Newer Evidence

By LEONARD WEISS AND JEFFREY G. WILLIAMSON*

The striking difference between the results reached by Charles Link, using published 1970 Census data, and ours, using the 1967 *SEO* tape, calls for some explanation. In an effort to find the source of the difference, this comment reports our results using the 1970 Census (1969 income and earnings) individual data from the public use tapes. We shall show that the source of the discrepancies lies with the form of the published Census data rather than in underlying economic relationships. Using individual data, we show that the effect of education on earnings is similar for blacks and whites in 1969, a result which confirms our assertions from the 1966 *SEO* data. We suspect that similar results would be forthcoming from the 1959 Census data. It seems likely that the low apparent return to black education found by Lester Thurow was also a statistical artifact. At the very least, we conclude that the use of grouped data for policy analysis of this sort is very suspect and that individual data should be used wherever possible.

The published census data is beset with numerous flaws. First, it presents data on income rather than earnings from work or self-employment. Income includes welfare transfers, unemployment insurance, and the like. As a result, the effect of education on income is apt to be weak at low levels of education compared with its effect on earnings. Second, the published data excludes persons with zero income. Since these include the unemployed, their exclusion tends to weaken the education-earnings relationship still further. Third, the published data include all males in the armed forces and in institutions (for example, jail). The education-earnings relationship may be biased downwards even further since income in kind is an enormous component of military pay. Each of these flaws in the published data apply to whites as well as blacks, but the distortion is likely to be more severe for blacks since they have higher dependence on transfers, higher unemployment rates, and higher participation in the armed forces.

In addition, the published Census data consist of median incomes for age-education groups, and three problems arise as a result. First, the census groups are quite broad. For example, median incomes are reported for those with 1–4 years of education, 5–7 years of education, etc., and for ages 18–19, 25–29, . . . , 36–44, etc. In contrast, the 1967 *SEO* tape and the 1970 Census public use tapes permit us to use precise age and education data. Second, the Thurow and Link studies gave each age-education cell equal weight, while our study is based on individual data effectively weighted by the number of individuals in a given cell. Third, the published cell medians do not reflect outliers very well, but they are included in our study. Obviously, a grouped data regression will always yield a higher R^2. Beyond that, it is not clear what a priori bias is introduced by grouping.

In an attempt to evaluate the effect of these differences we grouped the public use tape data following the census procedures, but utilizing alternative measures of income and alternative sample definitions. These results were then compared with those derived from individual data upon which the grouping was based. The results of these experiments appear in Table 1 for those in the North, Table 2 for those in the South, and Table 3 for the nation as a whole.

---

* The University of Wisconsin, Madison. We gratefully acknowledge the superb research assistance of Nancy Williamson. The research was supported by funds granted to the Institute for Research on Poverty at the University of Wisconsin by the Office of Economic Opportunity. The conclusions are the sole responsibility of the authors.

Table 1—Elasticities of Income and Earnings with Respect to Education and Experience: Northern Males, 18–64; 1969

Specification	Years of Education			Years of Experience			Number of Individuals Upon Which Regression Observations are Based
	0–8	9–12	13+	0–5	6–15	16+	
Northern Whites							
1. Link's results	.06	.09	2.32	.14	.84	−.18	
2. Same specifications using grouped data from public use tapes	.04	.39	2.04	.13	.83	−.15	334,317
3. Same as 2 but including zero incomes	.12	.58	1.99	.12	1.02	−.18	344,978
4. Same as 3 but using earnings, not income	.18	.63	1.95	.13	1.21	−.28	344,978
5. Same as 4 but excluding those in armed forces and institutions	.10	.74	1.82	.12	.89	−.23	299,390
6. Same as 5 but using individual data	1.11	1.56	1.68	.66	.88	−.77	290,632
Northern Blacks							
1. Link's results	.03	.46	1.74	.17	.67	−.25	
2. Same specifications using grouped data from public use tapes	.02	.57	1.82	.09	.87	−.30	23,046
3. Same as 2 but including zero incomes	.06	1.80	1.48	.16	1.42	−.44	25,222
4. Same as 3 but using earnings, not income	.11	1.74	1.50	.15	1.57	−.49	25,222
5. Same as 4 but excluding those in armed forces and institutions	.04	1.19	1.86	.16	1.07	−.34	21,735
6. Same as 5 but using individual data	.81	1.55	2.81	1.31	.82	−.73	20,913
7. Weiss-Williamson 1966 results	.21	1.46	1.51	1.70	.54	[a]	

[a] Our 1966 results included a fourth experience variable for those with 35 or more years experience. Thus, the estimated elasticity in this range is not comparable with the above regressions.

Table 2—Elasticities of Income and Earnings with Respect to Education and Experience: Southern Males, 18–64; 1969

Specification	Years of Education			Years of Experience			Number of Individuals Upon Which Regression Observations are Based
	0–8	9–12	13+	0–5	6–15	16+	
Southern Whites							
1. Link's results	.09	.62	2.16	.16	.69	−.17	
2. Same specifications using grouped data from public use tapes	.07	.66	2.11	.14	.73	−.16	133,994
3. Same as 2 but including zero incomes	.13	.81	2.21	.15	.94	−.19	139,474
4. Same as 3 but using earnings, not income	.19	.73	2.18	.14	1.10	−.35	139,474
5. Same as 4 but excluding those in armed forces and institutions	.14	.91	2.14	.12	.83	−.27	117,664
6. Same as 5 but using individual data	1.16	1.70	1.75	.73	.88	−1.19	113,684
Southern Blacks							
1. Link's results	.06	.62	1.93	.16	.55	−.25	
2. Same specifications using grouped data from public use tapes	.03	.64	2.06	.10	.64	−.28	24,078
3. Same as 2 but including zero incomes	.07	.73	2.43	.15	.87	−.31	26,492
4. Same as 3 but using earnings, not income	.09	.72	2.39	.17	.92	−.38	26,492
5. Same as 4 but excluding those in armed forces and institutions	.08	.87	1.98	.18	.67	−.32	22,701
6. Same as 5 but using individual data	.61	.76	2.82	1.09	.85	−1.23	21,257
7. Weiss-Williamson 1966 results	.19	1.66	2.22	1.51	.78	[a]	

[a] Our 1966 results included a fourth experience variable for those with 35 or more years experience. Thus, the estimated elasticity in this range is not comparable with the above regressions.

TABLE 3—ELASTICITIES OF INCOME AND EARNINGS WITH RESPECT TO EDUCATION AND EXPERIENCE: ALL MALES, 18–64; 1969

	Years of Education			Years of Experience			Number of Individuals Upon Which Regression Observations are Based
Specification	0–8	9–12	13+	0–5	6–15	16+	
			All Whites				
1. Link's results	.07	.42	2.16	.15	.77	−.18	
2. Same specifications using grouped data from public use tapes	.05	.52	2.09	.12	.80	−.19	468,311
3. Same as 2 but including zero incomes	.13	.73	2.08	.13	.97	−.17	484,452
4. Same as 3 but using earnings, not income	.19	.74	1.99	.13	1.14	−.27	484,452
5. Same as 4 but excluding those in armed forces or institutions	.11	.90	1.88	.12	.86	−.25	417,054
6. Same as 5 but using individual data	1.24	1.63	1.68	.68	.88	−.88	404,316
			All Blacks				
1. Link's results	.05	.79	1.77	.16	.62	−.25	
2. Same specifications using grouped data from public use tapes	.03	.83	1.84	.09	.73	−.28	47,124
3. Same as 2 but including zero incomes	.06	1.27	1.87	.15	1.00	−.34	51,714
4. Same as 3 but using earnings, not income	.09	1.29	1.91	.15	1.11	−.41	51,714
5. Same as 4 but excluding those in armed forces or institutions	.07	1.10	1.96	.13	.84	−.35	44,436
6. Same as 5 but using individual data	.69	1.23	2.76	1.18	.85	−.99	42,170
7. Weiss-Williamson results for 1966	.21	1.80	1.77	1.60	0.69	[a]	

[a] Our 1966 results included a fourth experience variable for those with 35 or more years experience. Thus, the estimated elasticity in this range is not comparable with the above regressions.

The first line in each of these tables reports Link's estimated elasticities. The second line shows the results derived from the public use tapes using the same population concept, the same income definition, and the same cell groupings as that published by the Census. Lines 1 and 2 differ very little, and the discrepancies are presumably due to sampling errors since our grouping procedure relied on a one in one-hundred sample.[1] Subsequent comparisons will be made with line 2 since all the remaining elasticity estimates in these tables are based on the public use tapes.

Line 3 reports elasticities when those with zero incomes are included. The result is to increase the impact of education on income in low education ranges for both races and in both parts of the country. Line 4 reports elasticities when we retain those with zero incomes, but we now also use cell median earnings from work or self-employment instead of income. Again the education elasticity in the 0–8 and 9–12 year ranges are consistently raised. Note also that northern blacks are more greatly affected than northern whites. Line 5 shows the results using the same procedures but excluding those in the armed forces or institutions. The results of this change are surprisingly equivocal.

Finally, line 6 shows the results for individual data using earnings of the noninstitutional population, and including those with zero incomes.[2] The shifts in the estimated elasticities when the individual data are used are often enormous. For low education ranges in particular, the elasticities increase by factors of five to ten. Our 1966 estimates based on the *SEO* tape are reproduced on line 7. The specifications for those are the same as in line 6. The differences between the 1966 (line 7) and 1969 (line 6) estimates require explanation, rather than comparisons between Link (line 1) and ourselves. Part of the explanation may be attributed to the fact that the *SEO* sample intentionally overrepresented the poor. The public use tapes do not overrepresent any group relative to its role in the population. One thing is certain—sampling errors cannot be large. There were 42,170 observations in the all black regression and 404,316 in the all white sample.[3]

---

[1] Another explanation for the slight discrepancies may be that we use whites and blacks while the Census tables used by Thurow and Link are for whites and nonwhites.

[2] Because the dependent variable is the *log* of earnings, those with zero earnings were assigned an income of $1. Those with negative earnings and those with zero education or experience were excluded from the sample. The largest group of exclusions were persons with zero experience (age 18 or in their first year out of school). Persons with earnings in excess of $50,000 were assigned the estimated mean earnings for their occupation class estimated on the assumption of a Pareto distribution (following Herman Miller).

It is not clear why the grouped data yield such drastically different results from individual data. One possible reason is that the Thurow-Link regressions are unweighted. However, a weighted regression for northern blacks cell median earnings for the noninstitutional population yielded elasticities similar to those in line 5 of Table 1 (.05, .72, and 1.73 for the three education ranges). The assignment of $1 to those with zero incomes in our individual observation regressions was, of course, arbitrary, but we feel it was conservative. If the unemployed had been assigned an income of 1 cent as Thurow did, we would expect the effect of education in the low range to be even stronger. In any event, when the black regressions were run again excluding those with zero income, a large though weakened effect for education remained. (For northern blacks the education elasticities were .16, .48, and 1.50. For southern blacks, they were .29, .64, and 1.66.) Finally, it should be emphasized that individuals with negative income are excluded from the individual regressions. This procedure was unavoidable given that the dependent variable is the *log* of earnings. Since only those with high permanent income (for example, high skill and education) can afford transitory negative income, it appears inevitable that the elasticities would be higher in the individual regressions. It also seems apparent to us that these individual elasticities are therefore more relevant for policy purposes.

An important conclusion from this exercise is that the effect of education on earnings is roughly as strong for blacks as for whites. We suspect the same was true in 1959 as well. The studies that yielded low returns for that period were based on grouped data like that used by Link. The widespread belief that education does not pay for blacks is certainly incorrect today Our results for 1969, like those we reached using 1966 data, still point to the usefulness of investments in black education in reducing interracial inequality.

REFERENCES

C. R. Link, "Black Education, Earnings, and Interregional Migration: A Comment and Some New Evidence," *Amer. Econ. Rev.*, Mar. 1975, *65*, 236–40.

H. Miller, *Income Distribution in the United States*, Washington 1966.

L. Thurow, *Poverty and Discrimination*, Washington 1969.

L. Weiss and J. G. Williamson, "Black Education, Earnings, and Interregional Migration: Some New Evidence," *Amer. Econ. Rev.*, June 1972, *62*, 372–83.

Office of Economic Opportunity, "Survey of Economic Opportunity," (*SEO*) conducted spring 1967, available on tape, Data Bank, Univ. Wisconsin.

U.S. Bureau of the Census, *U.S. Census of Population: 1970*, Subject Reports, *Educational Attainment*, Final Report PC(2)-5B, Washington 1973.

---

[3] Use of the individual observations, of course, reduced R^2. In the all white sample, R^2 underlying the elasticities reported in line 5 is .818 while in line 6 it is .104. The t-ratios on the education variables ranged from 9 to 63 in the individual regression; they ranged from 1.6 to 5 on the grouped data.

Theory of the Firm Facing Uncertain Demand: Comment

By Robert M. Korkie*

In a recent issue of this *Review*, Hayne Leland concludes that the risk-averse quantity setting firm will produce less under uncertainty conditions than it would under certainty, a finding that is consistent with that of other authors including John Lintner (1970). This conclusion is based in part on his assumed property of the stochastic demand curve, called the principle of increasing uncertainty (*PIU*).[1] This paper suggests that the intuitively supported *PIU* (increases in expected total revenue are accompanied by increasing risk) is a special and unusual case.

In what follows I adopt one of the behavioral modes specified by Leland and make the same equilibrium assumptions. It is demonstrated that the principle of increasing uncertainty does not hold for one of the most widely employed measures of risk in the economic and financial literature.

Leland introduces uncertainty through the implicit demand relation

$$(1) \qquad f(p, q, u) = 0$$

where u is a random variable with subjectively estimated probability density $dF(u)$. Let us deviate somewhat from this approach and introduce uncertainty in the parameters of a linear demand relation (henceforth called the random coefficients model) given implicitly by

$$(2) \qquad g(p, q, \tilde{a}, \tilde{b}) = 0$$

where $\tilde{a}$ and $\tilde{b}$ have subjectively estimated two-parameter density functions.

Consider the all equity taxless firm whose single security equilibrium price is given by the Sharpe-Lintner-Mossin single period capital asset pricing model.[2] That is,

$$(3) \qquad V_t = \frac{E(\tilde{x}_t) - \lambda_t Cov(\tilde{x}_t, \tilde{M}_t)}{(1 + R_{F_t})}$$

where

$V_t \equiv$ the beginning of period t firm market value,

$\tilde{x}_t \equiv$ the cash profits paid at the end of t,

$\tilde{M}_t \equiv$ the market portfolio's end of period t wealth prior to dividend disbursements,

$R_{F_t} \equiv$ the risk free borrowing and lending rate for period t,

$\lambda_t = \dfrac{E(\tilde{R}_{M_t}) - R_{F_t}}{\sigma^2_{RMt} M_{t-1}} \equiv$ the market price per unit of risk per dollar of beginning of period t market portfolio wealth M_{t-1},

$\tilde{R}_{M_t} \equiv$ the market portfolio's rate of return for period t,

$\sigma^2_{R_{M_t}} \equiv$ the variance of the market portfolio's rate of return in period t,

$Cov \equiv$ the covariance operator.

Now let us choose as the behavioral mode a firm whose *ex ante* control is quantity produced over the single planning period. *Ex post* to the realization of the outcome of the stochastic demand curve, the firm then adjusts price p_t to clear the market of the produced and demanded quantity q_t. The quantity setting firm seeks to

$$\underset{q_t}{\text{Maximize}}\ V_t,$$

where V_t is given by (3) and $\tilde{x}_t$ is given by

$$(4) \qquad \tilde{x}_t = \tilde{p}_t q_t - c(q_t) - F_t$$

where the explicit random coefficient demand is[3]

$$(5) \qquad \tilde{p}_t = \tilde{a}_t - \tilde{b}_t q_t$$

---

* Assistant professor, Faculty of Business Administration, University of Alberta.

[1] Lintner (1970, p. 255) also invokes this property in a portion of his paper.

[2] See William Sharpe, Lintner (1965), Jan Mossin. For problems of assuming a constant risk price (λ_t), see Eugene Fama.

[3] It can be shown that the random coefficients model is equivalent to a model with a single additive heteroscedastic error term.

where $c(q_t)$ is the variable cost and F_t is the fixed cost. Total revenue is

$$(6) \qquad T\tilde{R}_t = \tilde{a}_t q_t - \tilde{b}_t q_t^2$$

and expected total revenue is

$$(7) \qquad E(T\tilde{R}_t) = E(\tilde{a}_t) q_t - E(\tilde{b}_t) q_t^2$$

for selected values of q_t. The riskiness of total revenue is obtained as

$$(8) \quad Cov\,(T\tilde{R}_t, \tilde{M}_t) = q_t\, Cov\,(\tilde{a}_t, \tilde{M}_t) - q_t^2\, Cov\,(\tilde{b}_t, \tilde{M}_t)$$

To determine if the principle of increasing uncertainty holds, let us examine the values of q_t for which $E(T\tilde{R}_t)$ and $Cov\,(T\tilde{R}_t, \tilde{M}_t)$ reach maximums. First-order conditions for a maximum of expected total revenue yield

$$(9) \qquad q_t^* = \frac{E(\tilde{a}_t)}{2E(\tilde{b}_t)}$$

which occurs where the elasticity of expected demand equals -1. Similarly, a maximum of the risk of total revenue occurs at

$$(10) \qquad q_t^{**} = \frac{Cov\,(\tilde{a}_t, \tilde{M}_t)}{2\, Cov\,(\tilde{b}_t, \tilde{M}_t)}$$

which we assume is a positive quantity. Second-order conditions for a risk maximum require that

$$(11) \qquad -Cov\,(\tilde{b}_t, \tilde{M}_t) < 0$$

which is the most challenging assumption under which *PIU* is disproved. Hence, the only condition under which $E(T\tilde{R}_t)$ and $Cov\,(T\tilde{R}_t, \tilde{M}_t)$ reach a simulatneous maximum is if

$$(12) \qquad \frac{Cov\,(\tilde{a}_t, \tilde{M}_t)}{Cov\,(\tilde{b}_t, \tilde{M}_t)} = \frac{E(\tilde{a}_t)}{E(\tilde{b}_t)}$$

Barring this specialized case, for some values of q_t, $E(T\tilde{R}_t)$ is increasing and $Cov\,(T\tilde{R}_t, \tilde{M}_t)$ is decreasing or vice versa. Therefore, the principle of increasing uncertainty is violated.[4]

Leland suggests in a footnote that if *PIU* is violated at some level of output and a given demand outcome, a continuous demand curve implies that *PIU* is violated in a small neighborhood of the point. It is possible, as seen from this comment, that the neighborhood may be very large, and in fact may be as large as the normal range of *ex ante* operation of the control variable q_t.

In a situation where total revenue is a random variable and costs are certain, it may seem paradoxical that increased expected total revenue is accompanied by a decline in riskiness. However, the notion of a proportional tradeoff between risk and return is a phenomenon of equilibrium and firms searching for an optimal level of output are not in equilibrium. Furthermore, it can be shown that in a general utility maximization model, if one does not restrict the deviations of marginal revenue and profit from their respective means to be of the same sign, then the effect of uncertainty on planned firm output is inconclusive.[5] This ambiguous result for price setting firms is reached by Leland for reasons related to uncertain costs.

REFERENCES

E. Fama, "Perfect Competition and Optimal Production Decisions Under Uncertainty," *Bell J. Econ.*, Dec. 1972, *3*, 509–30.

R. Korkie, "The Theory of the Firm Facing Uncertain Demand and Cost with Bayesian Revision Over Two Periods," unpublished doctoral dissertation, Univ. Washington 1973.

H. E. Leland, "Theory of the Firm Facing Uncertain Demand," *Amer. Econ. Rev.*, June 1972, *62*, 278–91.

J. Lintner, "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *Rev. Econ. Statist.*, Feb. 1965, *47*, 13–37.

———, "The Impact of Uncertainty on the

---

[4] As Leland and Agnar Sandmo suggest, the *PIU* holds under perfect competition. This conclusion is applicable here since perfect competition implies $b_t = 0$, with certainty.

[5] A sufficient means of obtaining Leland's conclusions is to alter the random coefficients model to a linear demand model where the error term is homoscedastic. See my 1973 paper.

Traditional Theory of the Firm: Price Setting and Tax Shifting," in J. W. Markham and G. F. Papanek, eds., *Industrial Organization and Economic Development; Essays in Honor of E. S. Mason,* Boston 1970.

J. McCall, "Probabilistic Microeconomics," *Bell J. Econ.,* Autumn 1971, *2,* 403–33.

J. Mossin, "Equilibrium in a Capital Asset Market," *Econometrica,* Oct. 1966, *34,* 768–83.

A. Sandmo, "On the Theory of the Competitive Firm Under Price Uncertainty," *Amer. Econ. Rev.,* Mar. 1971, *61,* 65–75.

W. Sharpe, "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," *J. Finance,* Sept. 1964, *19,* 425–42.

Theory of the Firm Facing Uncertain Demand: Reply

By HAYNE E. LELAND*

Although complicated by extraneous elements,[1] Robert Korkie's criticism is based on the correct assertion that the linear stochastic demand curve

$$(1) \qquad p = \tilde{a} - \tilde{b}q$$

may not everywhere satisfy the principle of increasing uncertainty (*PIU*). Korkie concludes, perhaps with a modicum of exaggeration, that the *PIU* is a "special and unusual case."

I show below that many linear stochastic demand functions will satisfy the *PIU* for the range of outputs relevant to a profit oriented firm, even though these functions may not satisfy the *PIU* for outputs which firms will never consider. And I further suggest that the failure of the linear demand function to satisfy the *PIU* for some range of q is perhaps a better argument against the use of the linear form over that range than against the *PIU*—which Korkie admits has great intuitive appeal.

First, rewrite (1) in a form comparable to my original analysis:

$$(2) \qquad p = a(u) - b(u)q$$

where u is random. Recall our assumption that larger u's can be associated with higher prices (more favorable demands). Sufficient conditions for this to be satisfied for all q are

$$(3) \qquad a'(u) \geq 0, \qquad b'(u) \leq 0$$

with at least one inequality holding strictly. (Primes indicate derivatives.) Conditions (3) are necessary as well as sufficient, and therefore must always be satisfied if (2) introduces uncertainty either in a purely additive $[b'(u)=0]$ manner or in a purely multiplicative $[a'(u)=0]$ manner.

In the Appendix of my original paper, the *PIU* was shown to require that $\partial^2 TR(q, u)/\partial q \partial u$ have the same sign as $E[MR(q, u)]$, where $TR(q, u)$ and $MR(q, u)$ are total and marginal revenue, respectively. Using (2) and (3),

$$(4) \quad \partial^2 TR(q, u)/\partial q \partial u = a'(u) - 2b'(u)q > 0$$

Therefore, the linear stochastic demand curve satisfies the *PIU* for all q such that expected marginal revenue is positive.

From equations (22)–(24) of my original paper, we see that a risk-averse firm will have a smaller output under uncertainty if the *PIU* is satisfied at $q = q_c$, where q_c is the output of the firm under certainty. But at q_c, $E[MR(q_c, u)] = MC(q_c) > 0$. We conclude that the class of linear stochastic demand functions which satisfy (3) will exhibit the *PIU* for all relevant levels of q, and therefore firms facing such demand functions will produce less under uncertainty than under certainty.

My examination indicates that linear stochastic demand curves satisfying (3) *are* consistent with the *PIU* over the range of outputs relevant to analysis. Nonetheless, a methodological note might be in order. Korkie uses a common functional form (the linear demand function) to criticize an intuitively attractive behavioral principle (the *PIU*). One is reminded of a previous conflict—between the quadratic utility function and the Pratt-Arrow principle of diminishing absolute risk aversion. The outcome of that conflict was not the rejection of the principle, but rather of the quadratic utility function as a realistic description of tastes. Perhaps the real message of Korkie's paper is that the linear stochastic demand function should be used with caution.

REFERENCES

R. M. Korkie, "Theory of the Firm Facing Uncertain Demand: Comment," *Amer. Econ. Rev.*, Mar. 1975, *65*, 245–47.

H. E. Leland, "Theory of the Firm Facing Uncertain Demand," *Amer. Econ. Rev.*, June 1972, *62*, 278–91.

---

* School of Business Administration, University of California, Berkeley.

[1] Korkie uses *Cov* (*TR*, *M*) as a measure of riskiness. Since my analysis does not introduce securities markets, it is difficult to compare his measure with my own measure of risk, as exposited in the Appendix of my paper. If, however, the firm's revenue is distributed independently of all other firms, Korkie's measure of risk is proportional to the variance of total revenue, since the firm is itself a component of M. In turn, variance of total revenue is equivalent to my definition of risk if the degree of uncertainty is small, and moments of the distribution of total revenue higher than the second can be ignored.

Nontraded Goods, Factor Market Distortions, and the Gains from Trade: Comment

By MURRAY C. KEMP and EDWARD TOWER*

In a recent article in this *Review*, Raveendra Batra considers a small country producing two traded goods and one nontraded good. He shows that for such a country, an increase in an already positive import duty may raise welfare and has asserted that this finding is contrary to a "standard theorem derived by Kemp."

We do not know of any standard theorem which is placed in jeopardy by Batra's finding. In that part of Kemp's *Pure Theory* referred to by Batra, it is stated that tariff-restricted trade is superior to no trade whether or not there are nontraded goods; and recently Kemp and Henry Wan have provided a fairly careful demonstration of the proposition. However the Kemp-Wan proposition seems to have little to do with Batra's. If Batra had been able to show that a tariff increment which finally eliminates

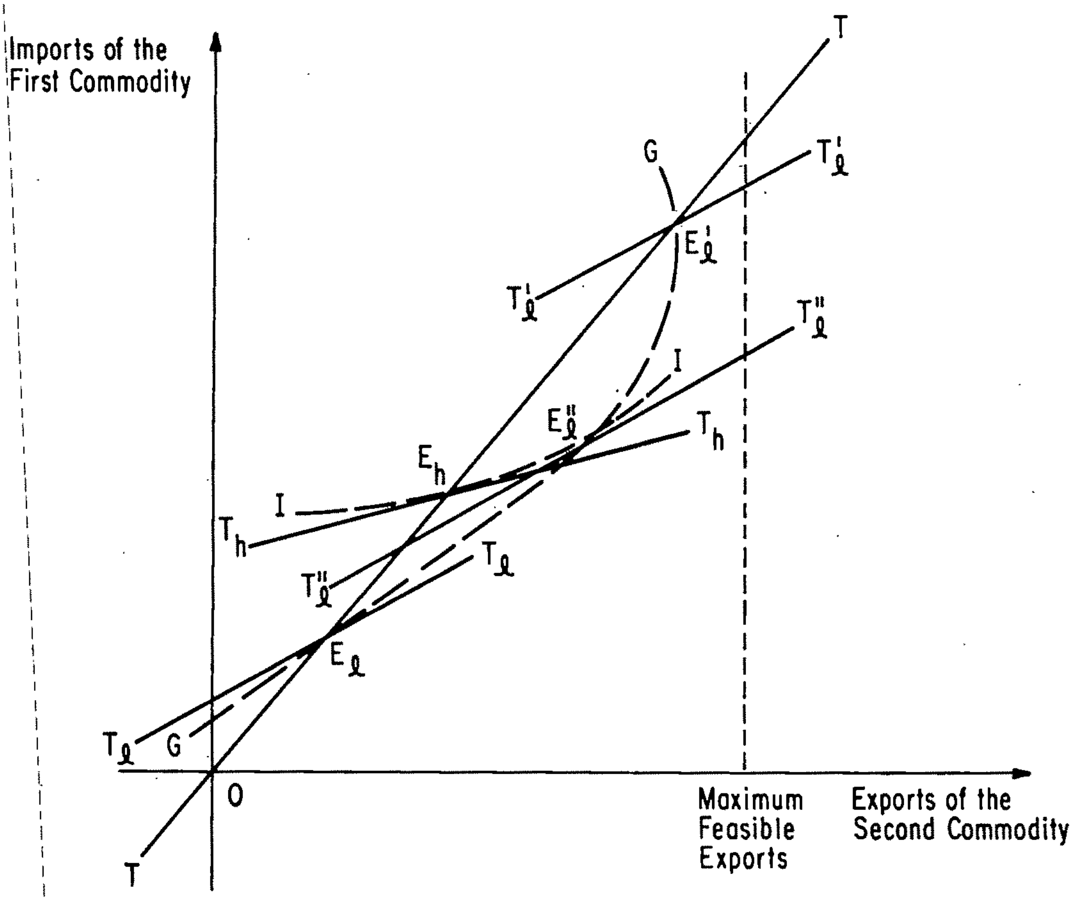


FIGURE 1

---

* Research professor of economics, University of New South Wales, and associate professor of economics, Duke University, respectively.

trade could raise welfare, then we should have agreed that the two propositions are in conflict. But this is precisely what Batra could not do. For from his equations (20) and (21) we obtain

$$\frac{1}{U_1}\frac{dU}{dt}=\frac{t}{1+t}\frac{dE_1}{dt}$$

and, if dt finally snuffs out trade, dE_1/dt is necessarily negative.

Moreover, the significance of Batra's finding must be assessed in the light of the fact that if there exists a tariff-distorted equilibrium with the property described by Batra, then there exists also an alternative preferred equilibrium without that property. That this is indeed a fact may be established by means of the accompanying trade indifference map. (Such a map can be drawn under any of the usual sufficient conditions, even when there are nontraded goods. The indifference loci have the usual curvature, and movements from one curve to another carry the usual welfare implications.) The given world terms of trade are indicated by the slope of the line TT; the lower-tariff domestic price ratio by the common slope of T_lT_l, $T_l'T_l'$, and $T_l''T_l''$; and the higher-tariff domestic price ratio by the slope of T_hT_h. The initial lower-tariff trade equilibrium lies, let us say, at E_l, where an indifference curve (not shown) is tangential to T_lT_l on TT. A new higher-tariff equilibrium lies, let us say, at E_h, where the indifference curve II is tangential to T_hT_h on TT. Evidently the community is better off at the higher-tariff equilibrium E_h than at the lower-tariff equilibrium E_l. Our diagram confirms Batra's algebra. Let us now consider the Engel curve GG defined by the lower-tariff domestic price ratio. This curve must pass through E_l and (given the curvature of II) through some point, say E_l'', on II to the right of E_h. After passing through E_l'', the curve may possibly continue to the right. However, it cannot cross the vertical line defined by maximum feasible exports. It therefore must eventually cross TT a second time, at say E_l'. Evidently, E_l' represents a second low-tariff equilibrium preferred to the first.

REFERENCES

R. N. Batra, "Nontraded Goods, Factor Market Distortions, and the Gains from Trade," *Amer. Econ. Rev.*, Sept. 1973, *63*, 706–13.

M. C. Kemp, *The Pure Theory of International Trade and Investment*, Englewood Cliffs 1969.

——— and H. Y. Wan, Jr., "The Gains from Free Trade," *Int. Econ. Rev.*, Oct. 1972, *13*, 509–22.

Nontraded Goods, Factor Market Distortions, and the Gains from Trade: Reply

By RAVEENDRA N. BATRA*

I take this opportunity to apologize to Murray Kemp for misinterpreting his theorem which was concerned with the comparison between restricted trade and no trade. I was wrong in thinking that this theorem examined the welfare implications of lower versus higher tariffs.

Coming now to the comments by Kemp and Edward Tower, I am afraid I do not find myself in complete agreement with them. In my paper I showed that in the presence of nontraded goods, a higher tariff will lead to a lower social welfare if the rise in tariff leads to a rise in the demand for imports, given, of course, that the country in question is small. Kemp and Tower, however, suggest that even if this is true, there exists an alternative low-tariff equilibrium with welfare level higher than that prevailing at the high tariff equilibrium.

I am not sure about the validity of this result because it depends on their assumption that the introduction of nontraded goods makes no difference to the properties of the trade indifference curve. I am not familiar with any study where the Kemp-Tower claim was established.

In what follows I show that although the trade indifference curve continues to be positively sloped even in the presence of the nontraded goods, its curvature need not be the same. First, I derive the equation of the trade indifference curve without nontraded goods and then show that if the social indifference curve is convex, then the underlying trade indifference curve is bowed-out towards the axis measuring exports.[1] In the presence of the nontraded goods, however, the convexity of the social indifference surface does not ensure the same curvature for the trade indifference curve.

Let C_i be the consumption of the ith traded good and X_i be its production, where $i=1, 2$. Assume that E_2 equals the amount exported of the second good, and E_1 is the amount imported of the first good. Then $C_1=X_1+E_1$, and $C_2=X_2-E_2$, so that a social utility function given by

$$U = U(C_1, C_2)$$

can be written as

$$U = U(X_1 + E_1, X_2 - E_2) \tag{1}$$

A trade indifference curve may be thought of a collection of points in the two dimensional E_2-E_1 space where utility is constant, and the marginal rate of substitution (U_1/U_2) equals the negative of the marginal rate of transformation (dX_2/dX_1), where $U_i=\partial U/\partial C_i$ is the marginal utility of the ith good and dX_2/dX_1 is the marginal rate of transformation. Thus, given that

$$\frac{U_1}{U_2} = -\frac{dX_2}{dX_1} \tag{2}$$

a utility level such as U_0 describes a trade indifference curve, where

$$U_0 = U(X_1 + E_1, X_2 - E_2) \tag{3}$$

The slope of U_0 is positive, because

$$dU_0 = U_1(dX_1 + dE_1) + U_2(dX_2 - dE_2) = 0$$

and in view of (2),

$$\frac{dE_2}{dE_1} = \frac{U_1}{U_2} \tag{4}$$

which is positive. The curvature of U_0 is defined by differentiating (4) with respect to E_1 and using (2) to obtain

---

* Professor of economics, Southern Methodist University.

[1] I am not aware of any study where the trade indifference curve has been mathematically derived.

$$(5)\ \frac{d^2E_2}{dE_1^2}=\frac{1}{U_2^3}[U_{11}U_2^2-2U_1U_2U_{12}+U_1^2U_{22}]\frac{dC_1}{dE_1}$$

where $U_{ij}=\partial U_i/\partial C_j\ (i\neq j)$. Now $dC_1/dE_1>0$ and the bracketed expression in (5) is negative because of the assumed convexity of the social indifference curve. Hence $d^2E_2/dE_1^2<0$ and it can be seen that the trade indifference curve will be bowed-out towards the exports axis, as in Figure 1 of the Kemp-Tower comment.

In the presence of the third nontraded good, where $C_3=X_3$ and the social utility function is

$$U = U(C_1, C_2, C_3)$$

the trade indifference curve is defined by

$$(6)\quad U_0 = U(X_1 + E_1, X_2 - E_2, X_3)$$

along with

$$(7)\quad U_1dX_1 + U_2dX_2 + U_3dX_3 = 0$$

Condition (7) is similar to condition (1) of the two-good case. The slope of the indifference curve as before is given by

$$dU_0 = U_1(dX_1 + dE_1) + U_2(dX_2 - dE_2) + U_3dX_3$$

so that in view of (7), $dE_2/dE_1=U_1/U_2$. Thus, in spite of nontraded goods, the trade indifference curve has the same positive slope as before. However,

$$(8)\quad \frac{d^2E_2}{dE_1^2} = \frac{1}{U_2^3}[U_{11}U_2^2 - 2U_1U_2U_{12} + U_{22}U_1^2]\frac{dC_1}{dE_1} + \frac{1}{U_2^3}[U_2^2U_{13} - U_2U_1U_{23} - U_2U_3U_{12} + U_3U_1U_{22}]\frac{dC_3}{dE_1}$$

Now the sign of dC_3/dE_1 and the bracketed expression of the second term in (8) is indeterminate. The sign of dC_3/dE_1, for example, depends on whether the nontraded good is a complement or substitute of the imported good. Hence the curvature of the trade indifference curve in the presence of the nontraded good is indeterminate.

REFERENCES

R. N. Batra, "Nontraded Goods, Factor Market Distortions, and the Gains from Trade," *Amer. Econ. Rev.*, Sept. 1973, *63*, 706–13.

M. C. Kemp and E. Tower, "Nontraded Goods, Factor Market Distortions, and the Gains from Trade: Comment," *Amer. Econ. Rev.*, June 1975, *65*, 249–50.

NOTES

The eighty-eighth annual meeting of the American Economic Association will be held in Dallas, Texas, December 28–30, 1975. The Employment Registry and Center will open December 27. The meeting was originally scheduled for October, but has been changed in the interest of attracting greater attendance. The preliminary program of the meeting will appear in the September 1975 issue of this *Review*.

In October 1974, the American Economic Association began publication of *Job Openings for Economists* (*JOE*), a bimonthly listing of job vacancies. The response from both employers and subscribers has been excellent. Employers may list vacancies free of charge and are encouraged to do so. The 1975 subscription rates are U.S.A., Canada, and Mexico (first class): $6.00 junior member, $12.00 others; foreign airmail: $12.00 junior member, $18.00 others. See the advertising section of this issue for a subscription form.

Economists who are *strongly* oriented toward the humanities, who use humanistic methods in their research, and who will be participating in meetings abroad that are concerned with the humanistic aspects of their discipline are eligible to apply for travel grants of the American Council of Learned Societies. Specifically, economists may be eligible if (a) they deal with the history of economic thought or economic history and (b) if their approach is qualitative and descriptive rather than quantitative and statistical. Conferences dealing with the establishment of social policy or legislation are generally ineligible. The ***deadlines for applications to be received*** in the office of the American Economic Association are: for meetings scheduled between June and September, *February 15;* for meetings scheduled between October and January, *June 15;* for meetings scheduled between February and May, *October 15.* Application forms may be obtained from Rendigs Fels, Secretary, American Economic Association, 1313 21st Avenue South, Nashville, Tennessee 37212.

Fulbright-Hays Program for Senior American and Foreign Scholars

Applications will be accepted this spring for more than 550 university lecturing and advanced research awards during 1976–77 in over 75 countries under the senior Fulbright-Hays program, the Council for International Exchange of Scholars announced recently. Specialists in economics and business administration who are U.S. citizens and have a doctorate or college teaching experience are invited to indicate their interest in an award by completing a simple registration form, available on request from the Senior Fulbright-Hays Program, 2101 Constitution Ave., N.W., Washington, D.C. 20418. Registrants will receive a detailed announcement of the 1976–77 program in April. The deadline for application is July 1, 1975 and it is also the suggested date for filing for lectureships.

Each year Fulbright-Hays agencies abroad forward to the Council applications of senior foreign scholars who are interested in remunerative appointments for lecturing and postdoctoral research at American colleges or universities for temporary periods. The scholars are eligible for Fulbright-Hays travel grants if arrangements are confirmed for remunerative appointments. The Council would be pleased to receive at any time information regarding appointments available at American educational institutions for foreign scholars for temporary periods of three months to one year. A directory of senior Fulbright-Hays foreign scholars who are in the United States this academic year is available on request to the Council.

An international conference on Regional Science, Energy, and Environment will be held at Centrum voor Economische Studiën, Katholieke Universiteit te Leuven on May 23–24, 1975 in cooperation with Regional Science Association (International) and its Dutch speaking section. Primary emphasis will be on the interrelationship of regional and interregional growth, energy shortages, and environmental repercussions. Professor Leo H. Klaassen, President of the Netherlands Economic Institute and another well-known scholar from the United States will be guest speakers. For further information, write to M. Chatterji, School of Management, SUNY-Binghamton, Binghamton, New York 13901 or P. Van Rompuy, C.E.S., K.U.L., Leuven, Belgium.

The International Institute of Public Finance will hold its yearly conference at Nice, France, Sept. 8–12, 1975. The conference theme is "Public Economics and Human Resources." The chairman of the program committee is Professor Victor Halberstadt, Leiden University, Netherlands.

A conference on "The Epistemological Relationship Between Science and the Humanities" will take place in the spring of 1976 at Miami University, Oxford, Ohio, with a planning conference expected to be held in the spring of 1975. The introductory question proposed is "Does intuition/perception/creativity have a similar function in the work of scientists and of humanists?" Those interested in participating may write for information to Professor Britton J. Harwood, Conference Secretary, Upham Hall, Miami University, Oxford, Ohio 45056.

The Association for the Study of the Grants Economy will be sponsoring a session on "The Grants Economy and Income Redistribution" at the 1975–76 Meetings of the Eastern Economics Association. Those interested in presenting papers on this topic should submit abstracts of no more than 300 words before June 1, 1975 to Professor Charles Waldauer, Widener College, Chester, Pennsylvania 19013.

The Economics Institute announces the 18th session of its summer program to be held in Boulder, Colorado, May 29–August 20, 1975. The program is designed for international students who are coming to the United States to undertake graduate studies in economics and related fields, including agricultural economics, management, and business. Participants will have an opportunity to review and strengthen their academic preparation and English proficiency through intensive instruction adjusted to their individual needs. The Institute is sponsored by the American Economic Association and is conducted in cooperation with admitting universities and student sponsoring agencies. For further information write: Director, Economics Institute, University of Colorado, Boulder, Colorado 80302.

New Scholarly Journal

The *Journal of Behavioral Economics* is an interdisciplinary journal aimed toward academicians, both inside and outside the economics discipline, and to practioners in business and nonprofit organizations. The two goals of the journal are to further knowledge of real world economic phenomena by integrating psychological and sociological variables into economic analysis, and to promote interdisciplinary research by academicians and practitioners dealing in economics, the behavioral sciences, and public policy. The editors are particularly interested in substantial articles which are coauthored by scholars representing different social science disciplines. Correspondence concerning subscriptions or contributions should be addressed to Dr. Richard E. Hattwick, Center for Business and Economic Research, College of Business, Western Illinois University, Macomb, Illinois 61455.

Research group at the Quebec Professions Board would be interested to contact any department or institution undertaking economic, sociological, and political research on professions and professionalism for exchange of information. Please write to Mr. Alain Vinet, Research Coordinator, Office des Professions du Quebec, 930 Chemin Ste-Foy, Suite 780, Quebec 6, P.Q., Canada.

The National Tax Association-Tax Institute of America announces the 1974 award winners in the annual competition for outstanding doctoral dissertations in government finance and taxation. The $1,500 first prize award was won by Helen F. Ladd of Harvard University (now at Wellesley College) with her entry, "Composition of the Property Tax Base and Public Expenditure Behavior." Two honorable mention awards of $500 each were won by William A. Fischel of Princeton University (now at Dartmouth College), "Fiscal and Environmental Considerations in the Location of Firms in Suburban Communities," and Ann R. Markusen of Michigan State University (now at the University of Colorado), "The Economics of Social Class and Metropolitan Local Government." The members of the 1974 Selection Committee were Professors Harvey E. Brazer, Arthur D. Lynn, Jr., Oliver Oldman, and James A. Papke. Information on the 1975 award competition may be obtained from Professor James A. Papke, Department of Economics, Krannert Graduate School of Industrial Administration, Purdue University, West Lafayette, Indiana 47906.

The Rockefeller and Ford Foundations Program in Support of Population Policy Research in the Social Sciences is continuing its worldwide program of awards designed to encourage population policy research on the social and economic determinants and consequences of population growth and distribution, and broaden the understanding of the relationship between population variables, law, and economic and social policies. The program has a July 1, 1975 deadline for submissions. Awards will be announced in December. The proposed research can begin as early as March 1, 1976. For further information outlining application procedures, write to The Rockefeller-Ford Program for Population Policy Research, 111 West 50th Street, New York, New York 10020.

Deaths

William R. Curtis, Bureau of Employment Security, U.S. Department of Labor (retired), Falls Church, Virginia, Oct. 13, 1974.

Stephen Enke, Manager, Washington Office, TEMPO, General Electric's Center for Advanced Studies, Sept. 21, 1974.

Benton P. Gillingham, graduate student, University of Wisconsin-Madison, Aug. 1974.

Seymour E. Harris, professor emeritus, department of economics, University of California, San Diego, Oct. 27, 1974.

Retirements

I. O. Linger, professor of economics, Texas A&M University, Sept. 1974.

Rosser B. Melton, professor of economics, North Texas State University, Aug. 1974.

Visiting Foreign Scholars

John R. Broome, Birkbeck College, England: visiting associate professor, department of economics, University of Virginia, Jan. 1975.

Josua W. Grobbelaar, University of South Africa: honorary fellow, University of Wisconsin-Madison, spring 1975.

Donald I. MacKay, University of Aberdeen: visiting professor of industrial and labor relations, School of

Industrial and Labor Relations, Cornell University, spring term, 1975.

Bengt Olov Muthén, University of Uppsala: honorary fellow, University of Wisconsin-Madison, 1974–75.

Toshio Ohtani, Aoyama Gakuin University: honorary fellow, University of Wisconsin-Madison, 1974–75.

Peter Zweifel, University of Zurich: honorary fellow, University of Wisconsin-Madison, 1974–75.

Promotions

Marcelle V. Arak: senior economist, Research and Statistics Function, Federal Reserve Bank of New York, Sept. 6, 1974.

Huntley H. Biggs: associate professor of economics, Colorado State University, July 1974.

Edgar N. Browning: associate professor of economics, University of Virginia, Sept. 1, 1974.

E. Gerald Corrigan: adviser and secretary, Secretary's Office, Federal Reserve Bank of New York, Sept. 6, 1974.

Suzanne Cutler: assistant secretary, Secretary's Office, Federal Reserve Bank of New York, Sept. 6, 1974.

Neville J. G. Doherty: associate professor, School of Dental Medicine, University of Connecticut-Farmington, Oct. 1, 1974.

John R. Dominguez: associate professor, business economics and finance, University of California-Los Angeles, July 1, 1974.

R. B. Ekelund: professor of economics, Texas A&M University, Sept. 1974.

James P. Gander: professor, department of economics, University of Utah, July 1974.

Martin A. Garrett, professor of economics, College of William and Mary, Sept. 1974.

Eva Garzoui: professor, department of economics, Ithaca College, Feb. 21, 1974.

William J. Gasser: chief, Balance of Payments Division, Federal Reserve Bank of New York, Sept. 5, 1974.

Victor P. Goldberg: associate professor of economics, University of California-Davis, July 1974.

W. P. Gramm: professor of economics, Texas A&M University, Sept. 1974.

Michael J. Hamburger: adviser, Research and Statistics Function, Federal Reserve Bank of New York, Sept. 6, 1974.

John A. James: assistant professor, department of economics, University of Virginia, Sept. 1, 1974.

William A. Jordan: professor of managerial economics, Faculty of Administrative Studies, York University, July 1, 1974.

Robert R. Keller: associate professor of economics, Colorado State University, July 1974.

Sima Lieberman: professor, department of economics, University of Utah, July 1974.

Philip W. Ljungdahl: professor of accounting, Texas A&M University, Sept. 1, 1974.

Bennett T. McCallum: professor, department of economics, University of Virginia, Sept. 1, 1974.

John R. McKean: professor of economics, Colorado State University, July 1974.

Emily P. Mackall: professor, department of economics, Youngstown State University, Sept. 1974.

Peter Miles: associate professor of economics, Sir George Williams University, June 1, 1974.

Michael P. Murray: assistant professor, department of economics, University of Virginia, Mar. 1, 1974.

James F. Niss: professor of economics, Western Illinois University, Sept. 1974.

Terutomo Ozawa: professor of economics, Colorado State University, July 1974.

Scott R. Pearson: associate professor, Food Research Institute, Stanford University, Sept. 1974.

Frederick C. Schadrack, Jr.: vice president, Bank Supervision and Relations Function, Federal Reserve Bank of New York, Sept. 6, 1974.

R. Elberton Smith: professor of economics, California Polytechnic State University.

C. Peter Timmer: associate professor, Food Research Institute, Stanford University, Sept. 1974.

R. D. Tollison: professor of economics, Texas A&M University, Sept. 1974.

Hylke Van de Wetering, Iowa-Peru Project; professor of economics, Iowa State University, Sept. 1, 1974.

Korman C. Walzer: associate professor of economics, Western Illinois University, Sept. 1974.

Administrative Appointments

Evangelos Djimopoulos: deputy chairman, department of economics and finance, Fairleigh Dickinson University.

J. Edwin Faris, Washington State University: professor and head, department of agricultural economics and rural sociology, Clemson University, Oct. 1, 1974.

Alexander Garcia: chairman, department of economics and finance, Fairleigh Dickinson University.

Martin J. Gerra: special assistant to the president, Gallaudet College, 1974–75.

George A. Hay: director of economics, Antitrust Division, U.S. Department of Justice.

Rodger E. Karrenbrock: chairman, department of graduate accounting, Woodbury University, Sept. 9, 1974.

Tassos G. Malliaris: chairman, department of economics, Loyola University of Chicago, Sept. 1, 1974.

Vinod P. Maniyar, chairman, department of economics, Christopher Newport College.

Henry G. Manne, University of Rochester: director, Center for Studies in Law and Economics, University of Miami, fall 1974.

Paul E. Merz: chairman, department of economics, St. Louis University.

Chris Musun: chairman, department of graduate business management, Woodbury University, Sept. 9, 1974.

Larry G. Pointer: head, department of accounting, Texas A&M University, Sept. 1, 1974.

Richard N. Rosett, University of Rochester: dean, Graduate School of Business, University of Chicago.

Tom S. Sale III: head, department of economics and finance, Louisiana Tech University, Sept. 1, 1974.

David L. Scott: chairman, department of business and economics, Florida Southern College, fall 1974.

Bruno Stein: director, Institute of Labor Relations, New York University.

Morton Stelcner: chairman, department of economics, Sir George Williams University, Mar. 1974.

Max D. Stewart: dean, School of Business and Economics, Wilfrid Laurier University, July 1, 1974.

R. D. Tollison: chairman, department of economics, Texas A&M University, Sept. 1974.

Keith K. Turner; chairman, economics department, University of Nebraska-Omaha, Sept. 1974.

Richard J. Ward, Peat, Marwick, Mitchell and Co.: dean, College of Business and Industry, Southeastern Massachusetts University, Dec. 1, 1974.

Robert C. Yost: chairman, department of graduate international business, Woodbury University, Sept. 9, 1974.

New Appointments

Trevor Bain, Queens College: professor, University of Alabama, Aug. 1974.

James J. Benjamin, Virginia Polytechnic Institute: associate professor of accounting, Texas A&M University, June 1, 1974.

Jeffrey I. Berkowitz, Ohio State University: Economic Policy Office, Antitrust Division, U.S. Department of Justice.

George C. Bitros: research associate, National Bureau of Economic Research, Sept. 20, 1974.

Dwight M. Blood: visiting professor, department of economics, Colorado State University, July 1974.

Jerry J. Bodily: Economic Policy Office, Antitrust Division, U.S. Department of Justice.

Gerald R. Bodisch: Economic Policy Office, Antitrust Division, U.S. Department of Justice.

Jon Breslaw: assistant professor of economics, Sir George Williams University, June 1, 1974.

James L. Butkiewicz: instructor, department of economics, University of Virginia, Sept. 1, 1974.

Dennis L. Chinn: assistant professor, Food Research Institute, Stanford University, Sept. 1974.

Thomas Chisholm, University of Missouri-Rolla: assistant professor, department of economics and business, Slippery Rock State College, Sept. 1, 1974.

Edward W. Clarke: lecturer, department of economics, School of Business and Economics, Wilfrid Laurier University, July 1, 1974.

Harold C. Cochrane: instructor, department of economics, Colorado State University, Sept. 1974.

Mark B. Cohen: Economic Policy Office, Antitrust Division, U.S. Department of Justice.

Kent A. Currie, University of Iowa: assistant professor of economics, Illinois Institute of Technology, Sept. 1974.

Phoebus J. Dhrymes: senior research associate, National Bureau of Economic Research, Sept. 16, 1974.

William M. Dugger, University of Texas-Austin: assistant professor, North Texas State University.

Rolf H. Dumke: lecturer, department of economics, School of Business and Economics, Wilfrid Laurier University, July 1, 1974.

Robert V. Eagly: economist, Foreign Research Division, Federal Reserve Bank of New York, Sept. 12, 1974.

William M. Edwards: instructor, department of economics, Iowa State University, July 1, 1974.

Charles M. Ellis, University of Alabama: North Texas State University, 1973–74.

Walter Enders: assistant professor of economics, Iowa State University, Sept. 1, 1974.

James A. Evenson: temporary assistant professor, department of economics, Colorado State University, Sept. 1974.

Norman Familant: instructor, department of economics, University of Virigina, Sept. 1, 1974.

Douglas Fisher: professor of economics, Sir George Williams University, June 1, 1974.

Mark B. Fitzsimmons: economist, Balance of Payments Division, Federal Reserve Bank of New York, Sept. 19, 1974.

Michael A. Flynn: instructor in economics, Wellesley College, Sept. 1974.

John Fox, Brown University: assistant professor, Youngstown State University, Sept. 1974.

Victor R. Fuchs: professor, department of economics, Stanford University, Sept. 1974.

Gasper A. Garofalo: assistant professor of economics, College of William and Mary, Sept. 1974.

Theodore A. Gebhard: Economic Policy Office, Antitrust Division, U.S. Department of Justice.

Lawrence J. Greenberg: economist, Balance of Payments Division, Federal Reserve Bank of New York, Aug. 1, 1974.

Peter R. Greenhalgh: Economic Policy Office, Antitrust Division, U.S. Department of Justice.

J. R. Hanson II: assistant professor of economics, Texas A&M University, Sept. 1974.

Barry C. Harris: Economic Policy Office, Antitrust Division, U.S. Department of Justice.

Marilyn Hastings: assistant professor of economics, Western Illinois University, Sept. 1974.

Wallace E. Huffman: assistant professor of economics, Iowa State University, Sept. 1, 1974.

Gloria J. Hurdle: assistant professor of economics, Colorado College, Sept. 1974.

Richard A. Ippolito: Economic Policy Office, Antitrust Division, U.S. Department of Justice.

Philip Jacobs: assistant professor of economics, Sir George Williams University, June 1, 1974.

Parameswar Krishnakumar, University of Florida: assistant professor, department of economics and business, Slippery Rock State College, Sept. 1, 1974.

Paul S. Lande: assistant professor of city and regional planning, Ohio State University, Sept. 1974.

Irving F. Leveson, New York City Health Services Administration: senior professional staff, The Hudson Institute, July 1974.

H. H. Macaulay, Jr.: visiting professor of economics, Texas A&M University, Sept. 1974.

Allan S. Mandel: assistant professor, Lyndon B. Johnson School of Public Affairs, University of Texas-Austin, Sept. 1973.

Stephen Marglin: visiting professor of economics, Sir George Williams University, June 1, 1974.

Brian Maris, Kansas State University: visiting assistant professor, department of economics and business, Slippery Rock State College, Sept. 1, 1974.

Curtis H. Martin: research associate, National Bureau of Economic Research, July 1, 1974.

Robert T. Masson, Northwestern University: Economic Policy Office, Antitrust Division, U.S. Department of Justice.

E. Scott Maynes, University of Minnesota: professor, department of consumer economics and public policy, Cornell University, Aug. 1974.

Stephen Mehay, visiting assistant professor, Sir George Williams University, June 1, 1974.

Frederick C. Meltzer: economist, Money and Finance Division, Federal Reserve Bank of New York, Sept. 19, 1974.

J. Meyer: assistant professor of economics, Texas A&M University, Sept. 1974.

Robert T. Michael: associate professor, department of economics, Stanford University, Mar. 1974.

Marc A. Miles, University of Chicago: instructor of economics, Loyola University of Chicago, Sept. 1, 1974.

Roger L. Miller, University of Washington-Seattle: professor of economics and associate director, Center for Studies in Law and Economics, University of Miami, fall 1974.

George R. Morrison, Graduate School of Business Administration, Rutgers University: visiting professor, department of economics, University of Viriginia, Jan. 14, 1975.

Frank Muller: associate professor of economics, Sir George Williams University, June 1, 1974.

William C. Myslinski: Economic Policy Office, Antitrust Division, U.S. Department of Justice.

Forest Nelson: research associate, National Bureau of Economic Research, June 1, 1974.

David C. Neveln: Economic Policy Office, Antitrust Division, U.S. Department of Justice.

David J. Nicol: visiting assistant professor, department of business administration, John Carroll University, 1974–75.

Fred Nold, Stanford University: assistant professor of economics, University of Santa Clara, Sept. 1974.

Carl W. O'Connor: assistant professor of economics Iowa State University, July 19, 1974.

Gerald P. O'Driscoll: assistant professor of economics, Iowa State University, Sept. 1, 1974.

John A. Otte: instructor and associate, department of economics, Iowa State University, Sept. 1, 1974.

Diann H. Painter, University of Illinois: assistant professor of economics, Wellesley College, Sept. 1974.

Carl J. Palash: economist, Money and Finance Division, Federal Reserve Bank of New York, Sept. 19, 1974.

Peter Pashigian: senior research associate, National Bureau of Economic Research, Sept. 1, 1974.

John S. Pettingill: acting assistant professor, department of economics, University of Virginia, Sept. 1, 1974.

James E. Post, St. Bonaventure University and State University of New York-Buffalo: assistant professor, School of Management, Boston University, Sept. 1974.

John H. Preston: Economic Policy Office, Antitrust Division, U.S. Department of Justice.

Gordon C. Rausser: professor of economics, Iowa State University, Aug. 15, 1974.

Charles F. Revier: instructor, department of economics, Colorado State University, Sept. 1974.

M. O. Reynolds: assistant professor of economics, Texas A&M University, Sept. 1974.

John E. Roemer, University of California-Berkeley: assistant professor of economics, University of California-Davis, July 1974.

Samuel Rosenberg, University of California-Berkeley: lecturer in economics, University of California-Davis, July 1974.

Stephen D. Silberman: Economic Policy Office, Antitrust Division, U.S. Department of Justice.

Edward W. Sparling: instructor, department of economics, Colorado State University.

Bruce R. Snapp: Economic Policy Office, Antitrust Division, U.S. Department of Justice.

J. Ronald Stanfield: assistant professor, department of economics, Colorado State University, Sept. 1974.

Keith G. Stanga, Louisiana State University: assistant professor of accounting, Texas A&M University, Sept. 1, 1974.

George L. Stansbury III, Naval Postgraduate School: instructor of business and economics, Florida Southern College, fall 1974.

Edith C. Suss, International Monetary Fund: lecturer in economics, University of Virginia, Jan. 14, 1975.

Rinaldo H. Toporovsky: research associate, National Bureau of Economic Research, Sept. 16, 1974.

James Townsend: associate professor of finance, Louisiana Tech University, Sept. 1, 1974.

Warren L. Trock: professor, department of economics, Colorado State University, Sept. 1974.

Hiroki Tsurumi: research associate, National Bureau of Economic Research, Sept. 16, 1974.

Kent Wall: research associate, National Bureau of Economic Research, Sept. 1, 1974.

Bernard L. Weinstein, Federal Trade Commission: senior research associate, Institute for Public Policy Alternatives, State University of New York, Dec. 1974.

Richard Weisskoff: associate professor of economics, Iowa State University, Sept. 1, 1974.

Patrick J. Welch: assistant professor of economics, St. Louis University.

Ronald Wienk: visiting instructor, department of economics and finance, Fairleigh Dickinson University, Sept. 1, 1974.

Sandra I. Williams: instructor, department of business and economics, Florida Southern College, fall 1974.

Dan Yaron: visiting professor, department of economics, Colorado State University, Sept. 1974.

Steven P. Zell: business economist, Federal Reserve Bank of Kansas City, Aug. 1974.

Leaves for Special Appointments

William Becker, Colorado College: visiting professor, University of the Americas, Mexico.

Huntley H. Biggs, Colorado State University:

development economist, Development and Resource Corporation of Sacramento, Jan.-Mar. 1975.

Gerald L. Childs, Rutgers University: Economics Policy Office, Antitrust Division, U.S. Department of Justice.

Edward C. Koziara, Drexel University: Fulbright lecturer in economics and industrial Relations, Royal University of Malta, Sept. 1974.

Karen S. Koziara, Temple University: Fulbright lecturer in economics and industrial relations, Royal University of Malta, Jan. 1974.

Robert H. McGuckin, University of California-Santa Barbara: Economic Policy Office, Antitrust Division, U.S. Department of Justice.

Donald L. Martin, University of Viriginia: research fellow, The Hoover Institution, Stanford University, 1974–75.

E. Scott Maynes, Cornell University: visiting research scientist, Survey Research Center, University of Michigan, 1974–75.

William C. Merrill, Iowa State University: agricultural economist, Agricultural Economics and Sector Planning Division, Bureau of Technical Assistance, USAID, Sept. 1, 1974.

J. C. Miller III, Texas A&M University: senior staff economist, Council of Economic Advisors, Sept. 1974.

Tridib K. Mukherjee, Appalachian State University: professor of agricultural economics and team member, Cameroon Project, Southern University, July 1, 1974.

Sung W. Son, Slippery Rock State College: vice president, Northwestern National Bank, July 1, 1974.

James A. Stephenson, Iowa State University: Macro Model Project Leader, AID-Thailand, Sept. 1, 1974.

Paul K. Worthington, Slippery Rock State College: Brookings Institution Economic Policy Fellow, Social Security Administration, July 1, 1974.

Ronald A. Wykstra, Colorado State University: economic adviser to government of Western Samoa, economic development department, Sept. 1974–Sept. 1975.

Resignations

David Cheng, Georgia Institute of Technology: University of Alabama, Sept. 1974.

James W. Christian, Iowa State University, June 1, 1974: vice president, National Savings and Loan League, Washington, D.C.

William P. Culbertson, University of Virginia, May 31, 1974: department of economics, Louisiana State University.

E. Carl Jones, Louisiana Tech University, Aug. 31, 1974: Mississippi State University.

Benton F. Massell, Food Research Institute, Stanford University: Office of Management and Budget, Executive Office of the President, Sept. 1974.

Tracy W. Murray, Georgia Institute of Technology: New York University, June 1974.

Erik Thorbecke, Iowa State University: Cornell University.

◆

NOTE TO DEPARTMENTAL SECRETARIES AND EXECUTIVE OFFICERS

When sending information to the *Review* for inclusion in the Notes Section, please use the following style:

A. Please use the following categories:

1—Deaths
2—Retirements
3—Foreign Scholars (visiting the USA or Canada)
4—Promotions
5—Administrative Appointments
6—New Appointments
7—Leaves for Special Appointments (NOT Sabbaticals)
8—Resignations
9—Miscellaneous

B. Please give the name of the individual (SMITH, John W.), his present place of employment or enrollment: his new title (if any), his next place of employment (if known or if changed), and the date at which the change will occur.

C. Type each item on a separate 3×5 card, and please do not send public relations releases.

D. The closing dates for each issue are as follows: *March*, November 1; *June*, February 1; *September*, May 1; *December*, August 1.

This announcement supersedes and replaces a letter which was sent annually from the managing editor's office. All items and information should be sent to the Assistant Editor, *American Economic Review*, Box Q, Brown University, Providence, Rhode Island 02912.

MAKE PLANS NOW

to attend the

1975 Meetings - Allied Social Science Associations

in

Dallas, Texas

Sunday, Monday, Tuesday,

December 28-30

Employment Center opens Saturday, December 27

This is a change from the October dates
previously announced.

Please note the 1976 ASSA meetings will be held in Atlantic City, September 16-18.

NEW

Caveat Emptor:
An Introductory Analysis of Consumer Problems
Roger M. Swagler, *Drake University*
February 1975 Paper 336 pages
Requiring no background in economics, this text applies fundamental principles of economic analysis to the actual here-and-now problems consumers face in the marketplace.

NEW

Money and Banking:
An Introduction to the Financial System
Thomas E. Van Dahm,
Carthage College
February 1975 Cloth [illegible] pages
[illegible]

NEW

Dialectical Economics:
An Introduction to Marxist Political Economy
Lyn Marcus, *Chairman, National Caucus of Labor Committees*
March 1975 Cloth 544 pages
An unprecedented approach to Marx's method and economic theory, this book explains, analyzes, and interprets Marxian economics through an interdisciplinary approach.
COMPLIMENTARY EXAMINATION COPIES NOT AVAILABLE.

NEW

Labor Market Segmentation
David M. Gordon,
New School for Social Research
Richard Edwards, *Harvard University*
Michael Reich, *Boston University*
January 1975 Paper 350 pages
This book examines persistent divisions within the labor force and proposes a theory of labor market segmentation to explain the continuing patterns of labor market differences.

The Theory of Microeconomic Policy
Nancy Smith Barrett,
The American University
1974 Cloth 384 pages
Rather than attempting to justify any particular economic system (capitalism, socialism, etc.), this intermediate-level text probes resource allocation per se, viewing optimal efficiency of allocation as a major economic goal.

Economics
Daniel R. Fusfeld,
University of Michigan
1972 Cloth [illegible] pages
[illegible]

HEATH

Call toll-free 800 225-1388.
In Massachusetts call collect 617-862-6650.
D.C. Heath and Company, 125 Spring Street, Lexington, Massachusetts 02173
D.C. Heath Canada, Ltd., Suite [illegible], [illegible] Street W., Toronto, Ontario [illegible]

"Public Enemy #1" – Inflation?

Is inflation really public enemy number one? Or is it merely the most superficial layer of a complex structure of interlocking problems?

Now the fourth edition of *The Economic Problem* offers your students the basic conceptual tools for answering questions like these and for understanding the functioning of the American economic system.

The Economic Problem

Robert L. Heilbroner
Lester C. Thurow

Always one of the most readable introductions to economics available, **The Economic Problem, 4th Edition** features some important changes that make economic theory even more accessible to your students:

- Lester Thurow, one of the nation's leading young economists, has joined forces with Robert Heilbroner. You can witness the benefit of their combined expertise in the strengthened section on microeconomics.
- The interdisciplinary orientation of **The Economic Problem** calls on history, philosophy, and sociology wherever appropriate to help trace the diversity of forces influencing the economic system.

- The long section on the historical background of modern economics has been condensed. Much of the material appears in boxed inserts in other chapters to add illustrative depth to the teaching material.

Also available as aids to student understanding are: a teacher's manual, transparencies, and a student workbook-reader compiled by the authors. 1975 704 pp. $13.95

For further information please write to: Robert Jordan, Dept. J-136, College Division, Prentice-Hall, Inc., Englewood Cliffs, N.J. 07632

Prentice-Hall

OXFORD

Microeconomics

The Analysis of Prices and Markets

DONALD DEWEY, Columbia University. An introduction to the fundamental theorems of modern microeconomic theory, this text is the first to deal rigorously with the multiplant enterprise and market concentration. It contains an exceptionally detailed treatment of the theory of the firm, capital and interest, and the pricing of factor inputs, and welfare economics. While the axiomatic character of modern microeconomic theory is firmly emphasized, its tools are applied to a wide range of policy problems that interest students: conservation of natural resources, regulation of public utilities, labor legislation, contracts of professional athletes, antitrust, and cost-benefit analysis of public projects.

1975 352 pp. 95 graphs $9.95

New volumes in the

Economic Development Series

General Editor: GERALD M. MEIER, Stanford University

Agriculture and Structural Transformation

Economic Strategies in Late-Developing Countries

BRUCE F. JOHNSTON, Stanford University, and PETER KILBY, Wesleyan University. In this volume, an agricultural economist and an industrial economist examine the problems of designing agricultural strategies which will help overcome the acute poverty that characterizes low-income nations. They analyze the interrelationships between agriculture and other sectors of the economy and consider the special problems and opportunities associated with late development. Historical chapters highlight the experiences of Japan, the U.S., Taiwan, Mexico, and the Soviet Union while discussion of contemporary developing countries focuses on the development and policy issues associated with the Green Revolution in India and Pakistan.

Spring 1975 480 pp. graphs and tables
cloth $11.95 paper $6.95

Employment, Technology, and Development

AMARTYA SEN, London School of Economics and Political Science. Focusing on the interrelationships between institutions, technology, and employment, the author investigates the institutional factors which affect employment policy making: modes of production, patterns of ownership, and systems of employment. The techniques of project evaluation and shadow pricing are discussed with particular reference not only to institutional features, but also to political feasibility and behavioral characteristics. Professor Sen illustrates his approach in four appendices which reflect India's experience in the field.

Spring 1975 175 pp. cloth, prob. $7.50 paper, prob. $3.50

In the following trilogy of policy volumes in international political economy, Professor Meier offers students a provocative approach to major policy problems involved in establishing a new international economic order. By blending specific source materials with a more general analytical narrative, the author immerses students in a set of actual policy-making situations so they can become concerned with the controversy of foreign economic policy analysis and discover criteria for judging the merits of alternative courses of action.

Problems of a World Monetary Order

GERALD M. MEIER. "An excellent up-to-date account of the theoretical issues and institutional changes. Meier's synthesis of excerpts from newspaper and official reports, congressional hearings and journal articles is very successful." —Asim Erdilek, Case Western Reserve University

1974 318 pp. paper $4.95

Problems of Cooperation for Development

GERALD M. MEIER. This book explores the ways in which governments, international organizations, and multinational enterprises can cooperate in reorienting development strategy toward more effective "outward looking" policies and policies of "redistribution through growth." The specific policy problems are approached from the standpoint of new strategies of aid diplomacy, resources diplomacy, and commercial diplomacy.

1974 272 pp. paper $4.95

Problems of Trade Policy

GERALD M. MEIER. "Fine idea that brings out complexity of issues well. . . ." —John Sheahan, Williams College. "A novel treatment, superbly executed." —R. W. Trenton, Oklahoma State University. "It has the touch of the 'real world' that students are anxious to absorb. And it has the right balance of economics and political decision making that makes it interesting. Well done."—Refugio I. Rochin, University of California, Davis

1973 304 pp. paper $4.95

International Economic Reform

Collected Papers of Emile Despres

EMILE DESPRES; edited by GERALD M. MEIER. "This collection of the papers of the late Emile Despres (most of which were previously unpublished and, thus, not accessible) is a valuable book which I can recommend to any serious student of international economics. . . . The book demonstrates Professor Emile Despres' genius as an extraordinary economist."—Norman C. Miller, University of Pittsburgh, in the *Journal of Economic Literature*

1973 311 pp. $7.50

OXFORD UNIVERSITY PRESS
200 MADISON AVENUE
NEW YORK, N.Y. 10016

D C HEATH AND COMPANY
125 Spring Street
Lexington, Massachusetts 02173

HEATH

Lexington Books

HEALTH MANPOWER AND PRODUCTIVITY

The Literature and Required Research
edited by John Rafferty, *U.S. Department of Health, Education and Welfare*

Federal, state and local government officials hope to check spiraling medical care costs in part, through increased manpower productivity in the health services sector. Does the potential for such increased productivity exist? In response, Rafferty critically reviews the research literature dealing with health manpower productivity, placing special emphasis on economic literature. He discusses directions for future research which seem most promising and catalogs the important questions that still need to be answered. Notes, tables, figures.

CONTENTS: **Part I. Issues in Health Manpower Productivity.** 1. Manpower Substitution in Ambulatory Care/Uwe R. Reinhardt and Kenneth R. Smith. 2. Productivity and Economies of Scale in Medical Practice/Richard M. Scheffler. 3. Effects of Incentives on Physician Performance/Frank A. Sloan. 4. Patient Participation and Productivity in the Medical Care Sector/Fredrick L. Golladay. 5. A Review of Productivity in Dentistry/Paul J. Feldstein. 6. Occupational Licensure and Health Care Productivity: The Issues and the Literature/H.E. Frech. **Part II. A General Overview.** 7. Research on Health Manpower Productivity: A General Overview/Jack Hadley.
240 pp. $15.00

To order direct, write Order Dept., Lexington Books, D.C. Heath and Company, 2700 North Richardt Avenue, Indianapolis, IND 46219

The Welfare State and Equality
Structural and Ideological Roots of Public Expenditures
Harold L. Wilensky
Conservatives complain that the welfare state punishes the diligent and rewards the idle. Radicals and others argue that it fails to reduce poverty, achieve equality, or promote social justice. In his newest work Wilensky taps the relevant social science studies and provides a readable, provocative guide through this maze of welfare rhetoric. 176 pages, $8.95

Swedish Economic Policy
Assar Lindbeck
Mr. Lindbeck has focused mainly on economic policy in Sweden since the end of World War II. He places considerable emphasis on experiments with new tools to stabilize business fluctuations. 239 pages, $12.00

Ivory and Slaves in East Central Africa
Changing Patterns of International Trade to the Late Nineteenth Century
Edward A. Alpers
In this history of a complex international trade system in East Central Africa, the author suggests that the historical roots of underdevelopment must be sought in this system which was initiated by the Arabs in the fifteenth century. 304 pages, $13.50

At your bookseller

University of California Press Berkeley 94720

Educated planning today leads to higher profits tomorrow.

1. ANALYSIS AND CONTROL OF DYNAMIC ECONOMIC SYSTEMS

By **Gregory C. Chow,** Princeton University

Presents a set of related techniques for analyzing the properties of dynamic stochastic models in economics and for applying these models in the determination of quantitative economic policy.

1975 336 pages (approx.) $19.95

2. LOCATIONAL DYNAMICS OF MANUFACTURING ACTIVITY

Edited by **Lyndhurst Collins,** University of Edinburgh and **David F. Walker,** University of Waterloo, Canada

Original essays interpreting changing patterns of industrial location in terms of historical and behavioral perspectives as well as contemporary theories and models.

1975 412 pages (approx.) Prob.$27.50

3. SOCIAL INFORMATION PROCESSING AND STATISTICAL SYSTEMS—CHANGE AND REFORM

By **Edgar S. Dunn, Jr.,** Resources for the Future, Inc., Washington, D.C.

This book analyzes the ways social data for management, public policy, and social science are acquired and processed and discusses the dilemmas this poses for those seeking to reconcile the need for individual freedom with the equally compelling need for social order.

1974 246 pages $12.50

4. SOCIAL INDICATORS AND SOCIAL THEORY
Elements of an Operational System

By **Karl A. Fox,** Iowa State University

Integrates theory, methods, and data across several social sciences, considering time budgets, behavioral setting, personality structure, occupational structures and many other areas to provide a rich theoretical perspective.

1974 328 pages $14.95

5. SPATIAL PERSPECTIVES ON INDUSTRIAL ORGANIZATION AND DECISION-MAKING

Edited by **F. E. Ian Hamilton,** University of London

Original research studies by well-known authorities focusing upon locational behavior by industrial organizations. Stresses types of industrial organization and decision-making behavior and their impact on industrial location patterns.

1975 533 pages $29.95

6. DOLLARS FOR NATIONAL GOALS
Looking Ahead to 1980

By **Leonard A. Lecht,** National Planning Association

Analyzes the changes in spending patterns during the 1960s, showing how these changes will influence a series of national goals and problems over the next decade.

1974 295 pages $17.95

7. STATISTICAL SAMPLING FOR AUDIT AND CONTROL

By **T. W. McRae,** The University of Bradford Management Centre

This book enables accountants to apply statistical sampling methods in auditing and financial control. It will help reduce the high set up and learning costs that statistical sampling incurs.

1974 279 pages $19.95

Available at your bookstore or mail this coupon to:
WILEY-INTERSCIENCE,
Dept. 526, Box 4569,
Grand Central Station,
New York, N.Y. 10017
In Canada: 22 Worcester Road, Rexdale, Ontario
A 4898-WI

Mail coupon to—
WILEY-INTERSCIENCE, Dept. 526, Box 4569,
Grand Central Station, New York, N.Y. 10017

Please send me the book(s) I have checked below to read and use free for 10 days. At the end of that time, if I am satisfied with my order and want to keep it, I will send you the amount indicated for each book received plus postage and handling. Otherwise, I will return the order and owe nothing.

☐ 1 (1 15616-7)
☐ 2 (1 16582-4)
☐ 3 (1 22747-1)
☐ 4 (1 27060-1)
☐ 5 (1 34715-9)
☐ 6 (1 52095-0)
☐ 7 (1 58991-8)

☐ My check (money order) for $______ is enclosed.
☐ Please bill me. (Restricted to continental U.S.)
☐ Please send me a list of local bookstores carrying your titles.

Name____________________

Affiliation____________________

City__________ State________ Zip________

Please add state and local taxes where applicable.
Prices subject to change without notice. 092 A 4898-WI

THE MOTIVATORS

CASEBOOK OF ECONOMIC PROBLEMS AND POLICIES: PRACTICE IN THINKING

By Rendigs Fels, Vanderbilt University,

and Robert G. Uhler, Federal Power Commission

176 pages, Paper, $4.50

"At last—a source book for open-ended questions"

That's what one educator had to say about this supplement for the principles of economics course. It's already in use by more than 125 colleges and universities across the country. It's a book that lays it on the line, challenges thoughtful responses from students concerning society's problems today: pollution; inflation; migrant workers; (even takes up ticket scalping). Uses the case method as pioneered by the Harvard Business School. As one of your colleagues sums it up: "An excellent way to integrate theory and practice in economics".

AN INTRODUCTION TO ECONOMIC ANALYSIS:

AN INNOVATIVE APPROACH TO THE STUDY OF ECONOMICS

By Donald W. Paden, University of Illinois

550 pages (tent.), Paper, $7.95 (tent.)

A self-contained package in one book.

Except for outside readings, here's a relatively self-contained text which includes the basic or core concepts of traditional economics analysis. Contains conventional text material, programmed material, lesson objectives, and test questions.

Emphasizes economic efficiency, employment, growth, inflation, money and foreign trade. Motivates students to participate from the standpoint of their own individual concepts and problems.

PRICES & MARKETS: A MICROECONOMIC TEXT

By Charles W. Baird, California State University, Hayward

400 pages (tent.), Cloth, $11.95 (tent.)

Emphasizes applications.

A new text for the junior/senior level. Presents microeconomic theory in a manner which convinces students it is just as applicable to real world phenomena as macroeconomics—even more applicable. Students are constantly motivated to ask about the usefulness of models used. Gives numerous applications of present value analyses such as investment in human and nonhuman capital, conservation, planned obsolescence, and predatory pricing.

Write for your complimentary examination copy. Please include course, number, average enrollment, and present text.

From the world of West
A new world of contemporary college publications

west's
college
department

MARKETING
50 W. Kellogg Blvd.,
St. Paul, Minn. 55102

WINTHROP PROUDLY ANNOUNCES...

Jan S. Hogendorn's
MODERN ECONOMICS: AN INTRODUCTION

A balanced core treatment of economics principles for one and two-semester courses.
Paperback, 475 pages, $7.95
Student Guide and Teacher's Manual available, by Cemal Yalinpala.

Douglas F. Dowd's
THE TWISTED DREAM:
CAPITALIST DEVELOPMENT
IN THE UNITED STATES SINCE 1776

A provocative component for Economics and History courses.
1975 336pp. paper $5.95

COMING IN FEBRUARY...

J. Walter Elliott's
MACROECONOMIC ANALYSIS

Macroeconomic theory presented at the intermediate level for application to business and economic decision-making.
500 pages, $12.95

WINTHROP PUBLISHERS, INC.
17 DUNSTER STREET
CAMBRIDGE, MASS. 02138

a Prentice-Hall company

WP

OUTSTANDING ECONOMICS TEXTS FOR PARTICULAR ECONOMICS COURSES

THE INTERNATIONAL ECONOMY, Fifth Edition

P. T. Ellsworth, formerly University of Wisconsin; and
J. Clark Leith, University of Western Ontario

Now in its fifth edition, this leading text continues to provide comprehensive coverage of the principles of international trade and finance from both the historical and theoretical perspectives. Thoroughly revised and updated, the book traces the development of an international economy, the theory and policies of international trade, finance and balance of payments problems, and the Eurodollar market. Diagrams and illustrations help the student visualize theoretical principles.

1975 approx 608 pages

MONEY, BANKING, AND THE ECONOMY, Third Edition

John A. Cochran, Arizona State University

This comprehensive text is designed to introduce the beginning student to the fundamentals of money and banking. It is particularly concerned with promoting an understanding of the public policy issues, and the problems and institutions that characterize the macroeconomic policies of the 1970's. The book achieves a balance between the theoretical and institutional approaches in its treatment of monetary and banking history, the money market, commercial banking, the Federal Reserve System, central banking, inflation, and the international monetary system. Emphasis throughout is on current problems and institutions, making this a complete, modern and intriguing introduction to money and banking.

1975 approx 640 pages

COMPARATIVE ECONOMIC SYSTEMS

Richard L. Carson, Carleton University

In this exciting and interesting presentation, COMPARATIVE ECONOMICS SYSTEMS focuses on the theoretical and analytic aspects of the different economic systems in the world today. By discussing comparative economic systems in terms of economic principles and concepts, the book explains general theory and describes the various national systems as they relate to the theory.

It first analyzes the universal economic problems of scarcity and distribution, then treats various economic solutions in terms of economic priorities. The way in which specific countries apply these economic solutions is described in conjunction with each solution. Numerous graphs, charts, and tables appear within the text.

1973 717 pages

MACMILLAN PUBLISHING CO., INC.
100A Brown Street, Riverside, New Jersey 08075

PRAEGER
SPECIAL STUDIES

MARKETING IN JAPAN
A Management Guide
M. Y. Yoshino
Foreword by *Patrick M. Boarman* and
Richard C. King
This practical analysis of consumer trends and market opportunities in Japan includes extensive surveys of major institutions and their functions, the Japanese distribution system, and recent market developments. Sponsored by the Center for International Business, Pepperdine University, Los Angeles.
176 pp. Jan., 1975 ISBN 0-275-09170-8 $15.00

CHINESE AND SOVIET AID TO AFRICA
Edited by *Warren Weinstein*
Surveys trends in Chinese and Soviet aid to Africa. Finds Chinese aid, more controlled than Soviet aid, is labor intensive, and therefore usually more relevant.
316 pp. Jan., 1975 ISBN 0-275-09050-7 $20.00

FOREIGN PRIVATE MANUFACTURING INVESTMENT AND MULTINATIONAL CORPORATIONS
An Annotated Bibliography
Sanjaya Lall
This annotated guide to the growing English-language literature on foreign private manufacturing investment and the multinational corporation includes both theoretical and applied studies. References mainly, but not exclusively, post-date 1965.
212 pp. Jan., 1975 ISBN 0-275-05570-1 $16.50

ECONOMIC GROWTH IN DEVELOPING COUNTRIES—MATERIAL AND HUMAN RESOURCES
Proceedings of the Seventh Rehovot Conference
Edited by *Yohanan Ramati*
Authorities discuss three essential development themes—resources, technology, and income distribution; external constraints on development; and planning and implementation. Published in cooperation with the Continuation Committee of the Rehovot Conference.
522 pp. Feb., 1975 ISBN 0-275-09300-x $25.00

THE FUTURE OF COMMERCIAL BANKING
Edited by *Wray O. Candilis*
Foreword by *Senator John Sparkman*
Examines commercial banking in relation to the public. Includes chapters on the changing structures of banking; bank holding companies; community affairs; automation and employment; competitive forces.
218 pp. Jan., 1975 TC 0940 ISBN 0-275-28760-2
$16.50

CAPITAL ACCUMULATION AND TECHNOLOGY TRANSFER
A Comparative Analysis of Nigerian Manufacturing Industries
D. Babatunde Thomas
Examines the relationship between capital formation and technological progress and explores its influence on value added per man in the manufacturing sector. Based on a case study of 800 Nigerian companies in 19 industries.
170 pp. Jan., 1975 TC 0960 ISBN 0-275-28771-8
$15.00

YUGOSLAV ECONOMIC DEVELOPMENT AND POLITICAL CHANGE
The Relationship between Economic Managers and Policy-Making Elites
Richard P. Farkas
This analysis of the present Yugoslav system's organizational and functional aspects links the growing power of the Enterprise Directors with those of the policy-makers, and places them into the context of national economic programs and international economic relations.
150 pp. Feb., 1975 ISBN 0-275-28780-7 $13.50

ECONOMIC GROWTH AND EMPLOYMENT IN VENEZUELA
An Analysis of an Oil-Based Economy
Mostafa F. Hassan
Emphasizes the need in developing countries for employment policies consciously aimed at attacking the unemployment problem. Explains the contributing factors in Venezuela's high rate of unemployment at a time when the economy is rapidly expanding.
204 pp. Jan., 1975 ISBN 0-275-09460-X $15.00

Please include TC and ISBN numbers when ordering
Order directly from praeger 111 Fourth Avenue, N.Y., N.Y. 10003

8 new economics texts–

HUTCHINSON **Money, Banking, and the United States Economy, 3rd Edition**

Harry D. Hutchinson, University of Delaware, has written Money, Banking, and the United States Economy, 3rd Edition, with today's student in mind. It is simple, readable, complete, and entirely rewritten to focus on the economic crises in the mid-1970's. Intended for use in a one-semester undergraduate course in the economics of money and banking, the book contains a broad range of monetary topics competently developed with a balanced mixture of institutional and theoretical content.
March 1975 approx. 560pp. $12.95

BARRETT **The Theory of Macroeconomic Policy, 2nd Edition**

This text shows students that macroeconomic policy must have a theoretical basis. Nancy Smith Barrett, American University, presents traditional macroeconomic theory in the context of a policy planning model. In this revision there is an expanded and updated treatment of the theory of inflation and its policy implications, with a chapter on the theory of wage and price controls. New material on unemployment and labor market policy has been added. More attention is focused on monetarism and other non-Keynesian approaches.
January 1975 approx. 480pp. $12.95

COHEN/CYERT **Theory of the Firm: Resource Allocation in a Market Economy, 2nd Edition**

Kalman J. Cohen, Duke University, and Richard M. Cyert, Carnegie-Mellon University, have designed this text as a concise statement of economic knowledge about the firm. It emphasizes the theoretical development from neoclassical theory to modern relations. Students going through the book will have comprehensive knowledge about the status of the economist's views on the firm.
February 1975 approx. 608pp. $13.95

ZAHN **Macroeconomic Theory and Policy**

Frank Zahn, Bowling Green State University, presents a new approach to intermediate macroeconomics — organized on the premise that once the basic theoretical and technical apparatus of an aggregate economic model is familiar to students, analysis of aggregate economic problems and considerations of policy alternatives can be undertaken with relative ease.
February 1975 approx. 320pp. $12.95

—from prentice-hall

ISARD **Introduction to Regional Science**

This text represents one of the first introductions to regional science ever written. Walter Isard, University of Pennsylvania, combines elementary material covering urban and regional economics, economic geography, city and regional planning, land use and environmental management, and relevant materials from other social sciences as they contribute to an understanding of how a city and region grows and develops.
January 1975 approx. 544pp. $11.95

SIMON **Applied Managerial Economics**

Julian L. Simon, University of Illinois presents the latest methods in decision analysis, but handles all problems with simple tabular methods that require no mathematics beyond arithmetic. Renders the theory of managerial economics operational and useful.
January 1975 approx. 560pp. $15.95

SNODGRASS/WALLACE **Agriculture, Economics, and Resource Management**

Milton M. Snodgrass, Mexico State University, and L.T. Wallace, University of California, outline a complete introduction to agricultural economics and agricultural business management.
January 1975 approx. 544pp. $13.95

TUCKER **Essentials of Economics**

James F. Tucker, Virginia Polytechnic Institute and State University, presents material providing not only a broad basis for better understanding of the American economic system, but also the fundamental tools needed for more effective decision-making regarding this system.
November 1974 368pp. paper $6.95

For further information on any of these texts write: Robert Jordan, Dept. J145, College Division, Prentice-Hall, Englewood Cliffs, N.J. 07632

prentice-hall

Lancashire in Decline

A Study in Entrepreneurship, Technology, and International Trade

By Lars G. Sandberg. An examination of the precipitous and disastrous decline, following World War I, of what has been described as the world's first modern industry and the major factor in the emergence of Great Britain as the world's leading industrial power, that looks beyond the customary charges of technological backwardness and entrepreneurial ineptitude usually invoked to explain the eventual collapse of the Lancashire cotton textile industry to essay an analysis, based on an impressive body of statistical data, of international developments in drastically altered economic circumstances and a rapidly changing world. Illus. $15.00

Environmental Quality and Society

Edited by Richard A. Tybout. A collection of original essays that examine the issues posed by the deterioration of the environment within a social context, on the assumption that it is through changes in our societal institutions that the growing threat to life will finally be reversed. The authors include Durward L. Allen, Arthur A. Atkisson, Jr., Robert Bierstedt, Paul F. Brandwein, Adolf Ciborowski, W. Wesley Eckenfelder, Jr., Morten Gorden, S. Smith Griswold, Orris C. Herfindahl, Charles L. Hosler, G. Evelyn Hutchinson, Justin W. Leonard, Michael McCloskey, Senator Gaylord Nelson, Edward Stainbrook, and Richard A. Tybout. $7.50

The Soviet Foreign Trade Monopoly

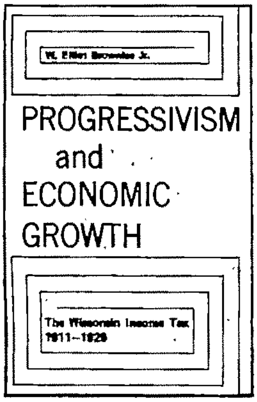
Institutions and Laws

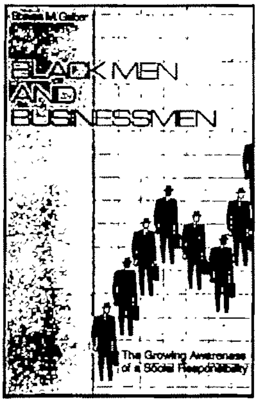
By John Quigley. An intensive study of the administration of what is the initial effort on the part of any government to conduct and to regulate, as a monopoly of the state, the foreign trade of an entire nation, which traces its evolution from its beginnings in 1918 in a decree signed by Lenin, who feared the effect on the fledgling socialist economy of capitalistic economic aggression from without, into the vast and complex bureaucracy that is the Soviet Trade Ministry of today. $15.00

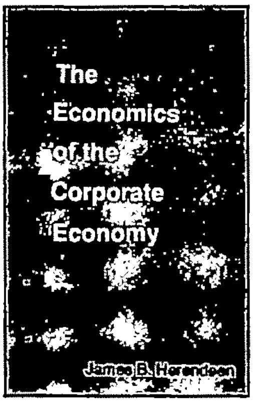
From the publishers of the *Journal of Money, Credit, and Banking.* Published quarterly in February, May, August, and November. Subscription rates: $14.00 per year; student subscriptions, $8.00 per year.

OHIO STATE UNIVERSITY PRESS
2070 Neil Avenue, Columbus 43210

New titles of interest


W. Elliot Brownlee Jr.
PROGRESSIVISM and ECONOMIC GROWTH
The Wisconsin Income Tax 1911–1929


Steven M. Gelber
BLACK MEN AND BUSINESSMEN
The Growing Awareness of a Social Responsibility



PROGRESSIVISM AND ECONOMIC GROWTH: The Wisconsin Income Tax–1911-1929
by W. Elliot Brownlee, Jr.
(1974) $10.00

BLACK MEN AND BUSINESSMEN: The Growing Awareness of a Social Responsibility
by Steven M. Gelber
(1974) $15.00

THE ECONOMICS OF THE CORPORATE ECONOMY
by James B. Herendeen
(Apr. 1975) $15.00

HOMAGE TO MALTHUS
by Jane Soames Nickerson, with an Introduction by Russell Kirk
(1975) $ 7.95

KENNIKAT PRESS
90 South Bayles Avenue, Port Washington, New York 11050

Economic Integration in Central America
Empirical Investigations
JEFFREY B. NUGENT

Nugent studies the experience of five countries in the Central American customs union in order to assess both the effects of the customs union itself and the potential benefits of supranational policy planning for developing countries. By constructing and analyzing smallscale econometric models of each country and the region as a whole, he finds that benefits from customs union participation are larger than what has generally been assumed. $14.00

The World between the Wars, 1919-1939
An Economist's View
JOSEPH S. DAVIS

"Absolutely magnificent! Such thoroughness such wit and wisdom. Nobody can pretend to know about those years until they have read this book." —*Samuel Ellot Morison*

The author here examines the fiscal policies in Europe and the U.S. at the time of the stock market boom and collapse and pinpoints those political and psychological misjudgements made by key individuals who helped to shape the depression's unfortunate course. $16.00

JOHNS HOPKINS
The Johns Hopkins University Press, Baltimore, Md. 21210

From The Free Press

THE UNCERTAIN SEARCH FOR ENVIRONMENTAL QUALITY

A Case Study in the Failure of Modern Policy Making

Bruce A. Ackerman and Susan Rose Ackerman, James W. Sawyer, Jr., Dale W. Henderson

"...an extraordinary study of the policy-making process. Better than any other analysis that I know of, it deals with the complex interaction among technical, economic, bureaucratic, political, and legal factors that are characteristic of the policy process....A landmark analysis which will be widely used for years to come...."—Henry S. Rowen, Professor of Public Management, Stanford University

367 pages $13.95

ECONOMIC FOUNDATIONS OF POLITICAL POWER

Randall Bartlett, Williams College

Refuting the traditional assumption that public policy is determined independent of conflicting private interests, this book examines the relationship between government and the economic agents of an industrialized society.

192 pages $8.95

MULTINATIONAL OIL

A Study in Industrial Dynamics

Neil H. Jacoby

Provides crucial information on how we got into the present energy crisis, then gives powerful alternative solutions. Jacoby presents new information on the foreign oil industry since World War II, including the development, structure, and performance of multinational corporations in the foreign oil market.

304 pages $3.95, paper

INCOME AND IDEOLOGY

An Analysis of the American Political Formula

Joan Huber and William Form, both, University of Illinois at Urbana-Champaign

Compares and contrasts the belief systems of poor, rich and middle-income groups, based on findings from a study of whites and blacks in a midwestern city and of Anglos and Chicanos in a rural Texas town. Areas which receive particular emphasis include: the causes of wealth and poverty, the importance of money, the "American Dream" of equal opportunity, and the extent to which Americans feel they can influence politics by voting. 226 pages $8.95

THE AMERICAN ECONOMY

An Historical Introduction to the Problems of the 1970's

Edited by Arthur M. Johnson, University of Maine at Orono

This book examines the quality of life, the morality of wealth, and the development of the American economy. Other contemporary issues presented in historical readings and explanatory essays include: poverty and the distribution of income; the tools used to manage and control our economy, their efficacy and relevance today; the role of private enterprise; and the prospects for the United States in a confused and changing world economy. 275 pages $4.95, paper

THE FREE PRESS

A DIVISION OF MACMILLAN PUBLISHING CO., INC.
100D Brown Street, Riverside, New Jersey 08075

The ECONOMICS INSTITUT

Announces its 18th Summer Program f

INTERNATIONA

Graduate Student

in ECONOMICS and RELATED FIELD

University of Colorado, May 29-August 20, 197

With sessions beginning May 29, June 16 (Main Session), and July 1

The Institute is specially designed for students from other countries w are planning to undertake studies at universities in the United States America in economics and related fields, including business and agricultur economics, during the subsequent academic year. Intensive instruction, a justed to individual student needs, is offered in each of the following co course areas:

- ECONOMIC THEORY
- ENGLISH
- MATHEMATICS
- STATISTI
- ECONOMIC PROBLEMS AND POLICIES
- BUSINESS ORGANIZATION AND MANAGEMENT

The primary purpose is to help students review and strengthen their academ preparation and English proficiency. However, the core course curriculu is supplemented by field trips and social and recreational activities design to acquaint them with campus and community life in the U.S.A.

The Institute is conducted in cooperation with admitting universities a sponsoring agencies. Proof of admission to an acceptable graduate progra in the following academic year is required.

INFORMATION: For further information and application materia write to: DIRECTOR, ECONOMICS INSTITUTE, UNIVERSITY COLORADO, BOULDER, COLORADO 80302, U.S.A.

Program sponsored by th

American Economic Associatio

and endorsed

The American Assembly of Collegiate Schools of Busine

The American Agricultural Economics Associati

Macroeconomics is more than just that...

Revised and up-dated!

MACROECONOMICS

SECOND EDITION

Norman F. Keiser, San Jose State University

The aim of **Macroeconomics** is not only to provide a thorough grounding in intermediate macroeconomic theory, but also to convey its significance in relation to current problems in American society.

Professor Keiser begins the text with a discussion of economic goals, the values they reflect, and the role of macroeconomics in attaining them.

Policy issues and empirical studies are integrated into the building of the macro model, making the material more realistic and relevant to students.

The second edition of **Macroeconomics** has been thoroughly up-dated and made more concise. The last chapter, "Fiscal and Monetary Policies From 1960-1974," is entirely new. Recent economic events have been incorporated into the text wherever relevant.

February, 1975 512 pages Order Code 31922

Accompanying **Macroeconomics** or any other standard macro text is **Macroeconomics: Problems and a Survey of Theory, Second Edition.** This workbook and brief review of macroeconomic theory has been revised and reorganized to match the text.

February, 1975 260 pages Order Code 31923

RANDOM HOUSE

The College Review Desk 400 Hahn Road, Westminster, Maryland 21157

INTRODUCTORY MACROECONOMICS

Peter C. Dooley, University of Saskatchewan

This basic introduction to macroeconomic theory:

- Begins with fundamental microeconomic concepts and principles as a foundation for developing macroeconomic theory.
- Provides thorough and concise treatment of the core of macroeconomic theory: consumption, investment, government, foreign trade, and money and interest.
- Relates economic policy to the interrelated problems of unemployment, inflation, and balance of payments.

Published 1974 192 pages Order Code 31757

ECONOMICS: MAINSTREAM READINGS AND RADICAL CRITIQUES

SECOND EDITION

Edited by **David Mermelstein,** Polytechnic Institute of Brooklyn

In this collection, articles by mainstream economists are juxtaposed with radical critiques. Students are presented with economic controversies and encouraged to reach their own conclusions.

Contributions are from prominent economists such as Milton Freidman and Paul Samuelson; radical economists such as Thomas Weisskopf and Paul Baran; and social commentators such as Ralph Nader, Robert Townsend, and William Domhoff. Topics include Sexism, Consumerism, Ecology, and the Industrial State.

Published 1973 484 pages Order Code 31731

Cornell
University
Press

New in paperback

The Participatory Economy

An Evolutionary Hypothesis and a Strategy for Development

By JAROSLAV VANEK. "This is a companion volume to Vanek's *General Theory of Labor-Managed Market Economies.* Here he handles the same theme in a simpler but broader way, describing how the growing influence of the workers leads to a gradual takeover of the management of producer units, and the impact of this on society.... This well-organized and clearly written work ... is valuable not only for students of economics, but also for those seeking a general knowledge of the social sciences."—*Library Journal* $2.45
(Also available in cloth/$10.50)

—Also by Jaroslav Vanek—

THE GENERAL THEORY OF LABOR-MANAGED MARKET ECONOMIES $16.50

MAXIMAL ECONOMIC GROWTH *A Geometric Approach to Von Neumann's Growth Theory and the Turnpike Theorem* $9.50

CORNELL UNIVERSITY PRESS ITHACA AND LONDON

A MODEL OF MACROECONOMIC ACTIVITY
Volume I: The Theoretical Model
Ray C. Fair, *The Cowles Foundation, Yale University*

Ray C. Fair develops a dynamic model of macroeconomic activity based on microeconomic foundations which accounts for all flows of funds in the system and for wealth and capital gains effects. This model overcomes many of the serious limitations of the standard Keynesian static-equilibrium model. Mr. Fair carefully presents the model in various stages to aid in understanding it and a number of important implications for the specification of empirical models are discussed. Volume II will present a relaxed empirical model.

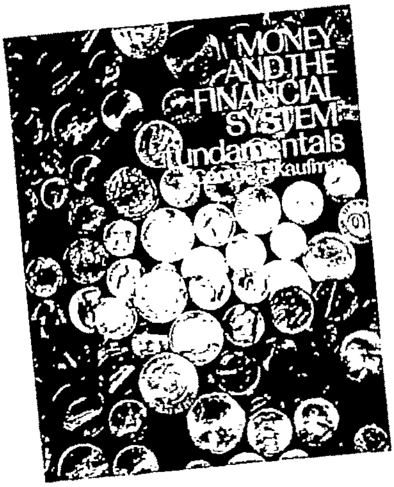
Contents:
1. Introduction. 2. Banks. 3. Firms. 4. Households. 5. The Government and the Bond Dealer. 6. The Dynamic Properties of the Model. 7. A Static-Equilibrium Version of the Model. 8. Conclusion.

260 pp. $13.50 Available

Ballinger
17 DUNSTER ST., HARVARD SQUARE,
CAMBRIDGE, MASSACHUSETTS 02138

NOW AVAILABLE!

A shorter, less demanding version of Kaufman's highly acclaimed MONEY, THE FINANCIAL SYSTEM, AND THE ECONOMY



MONEY AND THE FINANCIAL SYSTEM: FUNDAMENTALS

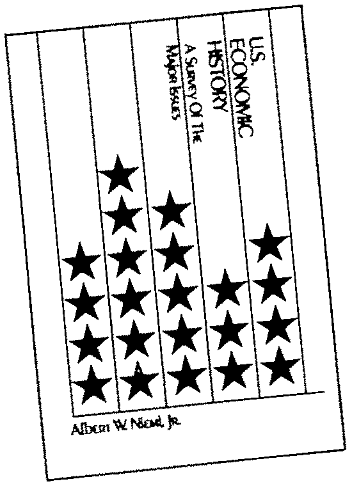
George G. Kaufman
University of Oregon

Covering all standard topics of introductory money and banking courses, this text clearly demonstrates how the financial sector is integrated into and interacts with the rest of the economy. It presents all the popular features of the author's earlier work but in an abbreviated format.

c. 425 pages/**$11.95 Cloth**
Instructor's Manual

NEW!

A successful integration of new and old economic research



U.S. ECONOMIC HISTORY: A Survey of the Major Issues

Albert W. Niemi, Jr.
University of Georgia

Written simply and concisely, this basic text is the first to combine the sophisticated quantitative-theoretical contributions of the new economic history and the rich stock of research generated by traditional economic historians so that both can be understood by students with no prior knowledge of economics.

c. 400 pages/**$7.95 Paper**

Rand McNally College Publishing Company
Box 7600/Chicago, Illinois 60680

Understanding Economics

Essays on Public Policy
edited by Yung-Ping Chen
University of California, Los Angeles

Eight original essays by **Lowell E. Gallaway** on unemployment; **Yung-Ping Chen** and **Charles T. Nisbet** on inflation; **David D. Martin** on the construction industry; **Abba P. Lerner** on exchange rate systems; **Pascal Salin** and **Sven W. Arndt** on devaluation; **Oswald H. Brownlee** on tax cuts; **Alan T. Peacock** on the performing arts; **Harry G. Johnson** on social issues.
Paper 151 pages 1974 $4.95

Little, Brown

Time on the Cross
The Economics of American Negro Slavery
Robert W. Fogel
University of Chicago, University of Rochester
Stanley L. Engerman
University of Rochester
Paper 286 pages 1974 $4.95
also available: a supplementary volume, *Evidence and Methods,* $6.95

The Economics of Industry
Roger Sherman
University of Virginia
Cloth 426 pages 1974 $12.95

National Monetary Policies and International Monetary Cooperation
Donald R. Hodgman
University of Illinois, Urbana
Cloth 266 pages 1974 $12.95

Taxation and Welfare
Arnold C. Harberger
University of Chicago
Paper 304 pages 1974 $6.95

International Trade and Finance
Readings
Robert E. Baldwin
J. David Richardson
both of University of Wisconsin, Madison
Paper 486 pages 1974 $6.95

World Trade and Payments
An Introduction
Richard E. Caves
Harvard University
Ronald W. Jones
University of Rochester
Cloth 574 pages 1973 $13.95

Family Finance
Money Management for the Consumer
Denis T. Raihall
American College
with the assistance of
Anthony D. Bilich
Cloth approx. 432 pages March 1975
$9.95 tent.
An Educational Associates Publication

Money Management for the Consumer
Readings and Cases
Denis T. Raihall
American College
with the assistance of
Anthony D. Bilich
Paper approx. 256 pages March 1975
$4.95 tent.
An Educational Associates Publication

an E . . .
is an E . . . is an
E . . .

No! An E is more than an E.

In the case of ECONOMICS — An Analysis of Principles and Policies, Second Edition, by Hailstones and Brennan, an E might well stand for Expectation, Enlightenment, Enrichment, Emphasis, and Explanation — for these are all hallmarks of this important new edition.

The real measure of a truly good textbook lies in its effectiveness as a teaching and learning tool, upon which the instructor can build an effective, interesting, and challenging fundamentals of economics course. Here are a few reasons why this new textbook merits your attention:

- the distinguished team of authors who are competent economists
- the appropriately balanced theoretical, graphic, and analytical treatment of economic fundamentals
- the unique blending of traditional theories and insights with contemporary economic issues
- the depth of coverage of macro-, micro-, and U.S. and world economics
- built-in flexibility through the wide variety of supplementary materials that correlate with the basic textbook

Request an examination copy of ECONOMICS — An Analysis of Principles and Policies, Second Edition, from the South-Western office serving you and you'll see why an E is more than just an E!

SOUTH-WESTERN PUBLISHING CO.
(Specialists in business and economic education)
Cincinnati, OH 45227
West Chicago, IL 60185
Dallas, TX 75229
Pelham Manor, NY 10803
Palo Alto, CA 94304
Brighton, England

Your colleagues comments on various chapters of

THE ECONOMY

By Bradley R. Schiller

2 "... exceptionally well written. The transition from micro to macro concepts is exceedingly well done. I particularly liked the repeated references to production possibility curves in developing the concepts."

—*Augustus Shackelford*
El Camino College

11 "A much better discussion than most others on the subject ... goes beyond mere description of the Phillips Curve to analysis of its origins and alternative policies."

—*Melville J. Ulmer*
University of Maryland

12 "... the chapter on energy and the environment is indicative of the blend of lively exposition and illustrations that ... characterize the book as a whole."

—*Wallace Oates*
Princeton University

14 "An excellent chapter, particularly on the weaknesses of the theory of diminishing marginal utility. This chapter not only brings basic economic theory to the student but helps him to analyze the faults of applying a scientific theory to human behavior."

—*Wesley Booth*
San Antonio College

15 "The micro races along, captures the students' interest and imagination."

—*Ernest Nadel*
University of California, Berkeley

20 "Schiller is very clear and convincing on horizontal monopoly/oligopoly power."

—*Walter Adams*
Michigan State University

25 "... clearly and concisely written. They manage to present the problems of poverty and welfare as analytic economic problems without losing a sense of the political and ideological concerns which make them issues at all ... the text has the important virtue of treating the reader with respect and speaking to him in a mature and intelligent tone."

—*Michael Piore*
Massachusetts Institute of Technology

26 "... contains one of the most realistic treatments of the problems of blacks and women in education and in the job market I have seen by an economist. It should give beginning students in economics exactly what they need to deepen their understanding of the economic and social issues at stake."

—*Barbara Bergman*
University of Maryland

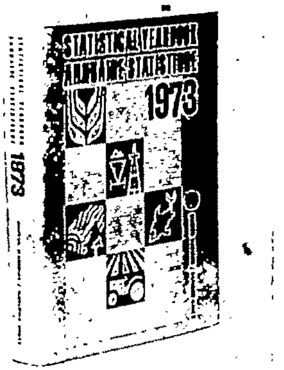
For further information, write to: Robert Jordan, Dept. J-139 College Division, Prentice-Hall, Englewood Cliffs, New Jersey 07632

January 1975, 864pp., cloth, $12.95

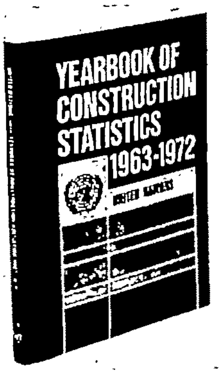
Prentice-Hall

PUBLICATIONS FROM UNITED NATIONS

STATISTICAL YEARBOOK 1973
Important compilation of statistics from countries throughout the world covering a wide range of economic and social subjects, including: population, agriculture, manufacturing, construction, transport, trade, balance of payments, national income, education and culture. Improved statistical coverage has enabled the *Yearbook* to widen the territorial scope of many of its tables and to provide more comprehensive and accurate world and continental aggregates.
Order No. E.74.XVII.1 Clothbound $35.00



COMPENDIUM OF HOUSING STATISTICS 1971
This world-wide compilation of housing and related statistics is prepared by the Statistical Office of the United Nations with the assistance of the Population Division and in co-operation with the Centre for Housing, Building and Planning. Its main purpose is to make available, information derived from national housing censuses and to provide for a more complete interpretation of this information than would otherwise be possible by presenting other data that have a bearing on the housing situation. These include data on population and its growth, dwelling construction, the cost of housing, and capital formation in housing.
Order No. E/F.73.XVII.4 $14.50



YEARBOOK OF CONSTRUCTION STATISTICS 1963-1972
This is the first in a new series of annual compilations of statistics on world construction activity and has been prepared to make available the information derived from annual surveys in this field undertaken all over the world.

The data presented in this publication are primarily based on replies to the Construction Statistics Questionnaire of the United Nations, distributed in 1972 for the first time. The information received has been supplemented in a few instances with data obtained from published sources, as well as with the work undertaken by the Statistical Office on the compilation of construction statistics in the United Nations *Statistical Yearbook*.
Order No. E.74.XVII.9 Clothbound $18.00

DEMOGRAPHIC YEARBOOK 1973
The twenty-fifth issue of the *Demographic Yearbook* contains 1973 statistics of area, population, natality, mortality, nuptiality and divorce for every country of the world, latest available data on expectation of life and a 7-year trend of international arrivals and departures. Data on the economic characteristics of the population are shown, including labour force participation rates by age and sex as well as various cross-classifications of population by industry, occupation, status, age and sex.
Order No. E/F.74.XIII.1 Clothbound $38.00

United Nations Publications, Room LX-2300, New York, N.Y. 10017
or
Palais des Nations, 1211 Geneva 10, Switzerland

• *New, Spring 1975*

GROWTH OF THE MODERN AMERICAN ECONOMY by STUART BRUCHEY, *Allan Nevins Professor of American Economic History, Columbia University*. Paperback, 160 pages, $3.95.

How a country with an agricultural economy moved to an industrial one in less than 200 years, and how it managed to raise the standard of living of its citizens to the highest in the world are the major themes of this concise overview. Questions related to continued growth are discussed.

SOCIAL WELFARE: A HISTORY OF THE AMERICAN RESPONSE TO NEED by JUNE AXINN and HERMAN LEVIN, both *University of Pennsylvania*. Paperback, 304 pages, $6.95.

A unique effort to assess the economic, social, political, and philosophical implications of the historical evolution of social welfare in America — supported by documents and illustrations.

DODD, MEAD & COMPANY, Inc., 79 Madison Ave., N.Y., N.Y. 10016

ECONOMICS

All NEW Titles For 1975

STATISTICAL INFERENCE FOR MANAGEMENT AND ECONOMICS
David V. Huntsberger, Iowa State University; Patrick Billingsley, University of Chicago; and D. James Croft, University of Utah. Instructor's Manual, Solutions Manual, and Workbook. 1975, Est. 544 pp.

FUNDAMENTAL STATISTICS FOR BUSINESS AND ECONOMICS
Abridged Fourth Edition
John Neter, University of Minnesota; William Wasserman, Syracuse University; and G. A. Whitmore, McGill University. Instructor's Manual, Solutions Manual, and Workbook. 1975, Paperbound, Est. 500 pp.

MACROECONOMICS
Peter E. Kennedy, Simon Fraser University
1975, Paperbound, Est. 480 pp.

URBAN AND REGIONAL ECONOMICS: Structure and Change
M. Jarvin Emerson, Kansas State University and Charles Lamphear, University of Nebraska. 1975, Est. 375 pp.

ALLYN AND BACON, INC., College Division, Department 893
470 Atlantic Avenue, Boston, MA 02210

The Second Edition of

ECONOMICS

AN INTRODUCTION TO TRADITIONAL AND RADICAL VIEWS

E. K. Hunt and Howard J. Sherman

This comprehensive introduction to economics, the first to compare traditional and radical views in a clear exposition for beginning students, has been updated and substantially revised in response to suggestions made by first edition users. Material on such topics as pure competition and monopoly, monopoly power, government and inequality, the economics of discrimination, and business cycles and unemployment has been thoroughly rewritten. New material has been added on aggregate economic problems and the institutional setting, consumption, wage-price controls, and waste and pollution. February 1975. Tentative: 672 pages $8.95/paper. Instructor's Manual. Part I will again be released in a separate volume. **PROPERTY AND PROPHETS: The Evolution of Economic Institutions and Ideologies, Second Edition:** February 1975. Tentative 192 pages; $3.95/paper.

Also New in 1974-1975

INTRODUCTION TO ECONOMETRICS: Principles and Applications

Harry H. Kelejian and Wallace E. Oates/September 1974. 300 pages; $11.95.

PROBABILITY AND MATHEMATICAL STATISTICS

Lester D. Taylor/October 1974. 346 pages; $14.95. Study Guide by Kushler: January 1975. $4.95 (tentative).

MONEY AND BANKING: Theory, Policy, and Institutions

Helen B. O'Bannon, David E. Bond, and Ronald A. Shearer/
February 1975. Tentative: 606 pages; $13.95. Instructors Manual.

INTRODUCTION TO MONETARY ECONOMICS

Stanley M. Besen/January 1975. 226 pages; $13.95.

URBAN ECONOMIC PROBLEMS

Richard F. Muth/January 1975. 402 pages; $12.95.

THE WOMEN'S MOVEMENT: Political, Socioeconomic, and Psychological Issues

Barbara Deckard/March 1975. Tentative: 480 pages; $6.95/paper.

Harper & Row
1817 10 East 53d Street, New York 10022

Leading books from RONALD

Weimer-Hoyt-Bloom: REAL ESTATE

Arthur M. Weimer, Indiana University; **Homer Hoyt**, Homer Hoyt Associates; and **George F. Bloom**, Indiana University. Selected by nearly 200 colleges and universities, this highly successful text covers the entire spectrum of today's real estate operations. Especially notable are the treatment of real estate decisions and decision makers, and the revised discussion of appraising methods. 6th Ed., 1972 831 pp. $13.50

Wren: THE EVOLUTION OF MANAGEMENT THOUGHT

Daniel A. Wren, Florida State University. ". . . the most intensive and encompassing treatment of the history of management and administration thought in existence today. No other single volume has attempted to trace the comprehensive story of management's emergence as a discipline from medieval times to the present proliferation of management theory."—BUSINESS HORIZONS. 1972 556 pp. $10.95

Braverman-Stewart: STATISTICS FOR BUSINESS AND ECONOMICS

Jerome D. Braverman and **William C. Stewart**, both, Temple University. Carefully selected topical coverage provides a lucid introduction to the principles of statistical inference as applied to business and economics. With an unusually large amount of attention paid to clear, concise explanations and illustrations, the concepts and methods of analysis are fully explored. 1973 482 pp. $11.95

Donaldson-Pfahl: PERSONAL FINANCE

Elvin F. Donaldson and **John K. Pfahl**, both, Ohio State University. Recognized as an outstandingly teachable book for more than two decades, this volume is designed to help students improve their understanding of planning, spending, saving, and investing. Major emphasis is placed on establishing and maintaining short- and long-term financial goals. Many illustrations. 5th Ed., 1971 802 pp. $11.50

Send for your copy of the new Ronald Catalog

THE RONALD PRESS COMPANY
79 Madison Avenue, New York, N.Y. 10016

SIMPLY OUTSTANDING

ECONOMICS AND SOCIETY

By Leonard W. Weiss,
University of Wisconsin, Madison

Simplicity. It's a guiding principle in Weiss. Each chapter of ECONOMICS AND SOCIETY opens with an explanatory paragraph . . . and closes with a summary. In between, your students will discover—with accumulating interest—economics as it has never before been introduced. Weiss is built upon a sound economic base. Yet its theory remains as simple as possible.

Most of it concerns supply and demand, or aggregate supply and demand. These concepts are applied repeatedly to such policy issues as **monopoly, discrimination, labor market, balance of payments,** and **public goods.**

Some texts are long, or diffuse, or incomplete for a one-quarter or one-semester introductory economics course, Weiss is **just right.** And Weiss is one author who knows how to express himself. Students will take to his lucid style at once. You have to read it to believe it.

Discover Weiss. You'll find that the outstanding coverage of ECONOMICS AND SOCIETY is structured with care around a single idea: Simplicity. January 1975 544 pages (tent.) $10.95

MACROECONOMICS:
Measurement, Theory, and Policy

By Michael C. Lovell, **Wesleyan University**

Lovell **is** macroeconomics. Designed for intermediate and upper-level courses, he stresses modern theoretical advances while analyzing unemployment, inflation, balance of payments, disequilibrium, business cycles, growth, and more.

The text is structured for maximum flexibility—with core chapters designated, key concepts and exercises at the end of chapters and appendices for more in-depth coverage of specific topics. All in all, Lovell provides a solid presentation of macroeconomic concepts. Spring 1975 550 pages (tent.) $13.95 (tent.)

THE ECONOMICS OF INTERNATIONAL BUSINESS

By R. Hal Mason, **UCLA,** Robert R. Miller,
University of Iowa, Dale R. Weigel,
University of Iowa and **International Finance Corp.**

Mason/Miller/Weigel shows how the tools of economic analysis can be applied to the problems faced by managers of international business. It's the only text • to discuss technology selection, and financial decision in the international firm • with an up-to-date treatment of international monetary affairs • which approaches international business from a moderately analytical viewpoint.
1975 444 pages $14.95

THE ECONOMICS OF ENVIRONMENTAL POLICY

By A. Myrick Freeman III, **Bowdoin College,**
Robert H. Haveman, **University of Wisconsin,**
and Allen V. Kneese,
Resources for the Future, Inc.

This brief text presents the facts of environmental problems, places problems in an economic framework, and stresses the attainment of effective environmental policy through an economic incentive strategy.
1973 184 pages $4.25-paper

For your complimentary examination copies, contact your Wiley representative or write to Mr. Art Beck, Dept. 498, N.Y. office. Please include course title, enrollment, and present text.

JOHN WILEY AND SONS, INC.
605 Third Avenue, New York, N.Y. 10016
In Canada: 22 Worcester Road, Rexdale, Ontario
Prices subject to change without notice.

A4879-PJW

wiley

New editions that reflect the new economic realities

INTERNATIONAL ECONOMICS

A Policy Approach

SECOND EDITION

Mordechai E. Kreinin, *Michigan State University*

As up-to-date as the current oil crisis, this new Second Edition of *International Economics* provides a comprehensive, truly international introduction to world finance and trade. Professor Kreinin emphasizes policy and the analysis of general principles, using the most important trading nation in the world—the United States—as his reference point for evaluating policy alternatives. The policy materials in this second edition reflect important developments in international economics through mid-1974—among them, the trade problems of developing countries, the 1974 GSP of the European Communities, the proposed GSP of the United States, the new international currency system and Euro-dollars, the balance of payment adjustments under fluctuating exchange rates, and the monetary approach to devaluation. As in the previous edition, the book thoroughly covers present currency arrangements, analyzes the balance-of-payments problem, offers suggestions for reforms of the monetary system, and explains the principles governing world trade, the role of trade organizations, and selected trade problems among developing nations.

500 pages

. . . and the latest economic thought

THE EVOLUTION OF ECONOMIC THOUGHT

THIRD EDITION

Jacob Oser and William C. Blanchfield, *Utica College of Syracuse University*

Now in its Third Edition, this readable, comprehensive survey of economic theory covers 72 economic theorists and their ideas from the 16th century to the present; it provides biographical sketches of each theorist as well as substantial quotations from their work and background material on the thought of each age. The authors evaluate each major school in terms of its social milieu, its essential ideas, and its impact on its own time and later periods.

Major changes, additions, and improvements in this Third Edition include a greater emphasis on the interrelationships among the different schools of economic thought; revised and expanded sections on Marx, Engels, and the Utopians; a revised section on monopolies; and extensive discussion of Galbraith's latest works, *The New Industrial State* and *Economics and the Public Purpose*; new treatment of our changed perception of economic problems as a result of the Keynesian revolution; and a new section on post-Keynesian economics that explores the contributions of Samuelson, Arrow, and the school of mathematical economics.

544 pages

HARCOURT BRACE JOVANOVICH, INC.
New York · Chicago · San Francisco · Atlanta

look at our statistics for 90 days . . . FREE

☐ **INTRODUCTORY STATISTICAL ANALYSIS**

By Donald L. Hamett, *Indiana University*, and James L. Murphy, *University of North Carolina* 1975 575 pages 103 illustrations

Designed for a beginning course in statistics for business, economics, and social science students, this text provides a mixture of intuitive explanation, examples, and mathematical rigor for students who have had exposure to college mathematics. Throughout the writing and classroom use of this text, careful attention has been given to readability, organization, and flow of material.

Numerous examples and problems show the usefulness of statistics in everyday decision making situations. Each chapter ends with a "Review Problems" section which reinforces the chapter material and an "Exercises" section which is more challenging. An Instructor's Manual is available.

☐ **INTRODUCTION TO STATISTICAL METHODS, SECOND EDITION**

By Donald L. Hamett, *Indiana University* 1975 640 pages 108 illustrations

The carefully selected level of presentation in this text allows students to make ample use of prior courses in college algebra, finite math, and calculus. A broad range of statistical topics are covered with the basic methods and applications for statistical analysis in business, economics, and most social science disciplines. An Instructor's Manual is available.

☐ **APPLIED STATISTICS FOR BUSINESS AND ECONOMICS**

By Henrick J. Malik and Kenneth Mullen, *University of Guelph, Ontario* 1975 565 pages 161 illustrations

The emphasis is on clarity and simplicity in this elementary text for students in business and economics with a high school algebra background. All concepts are introduced by examples, followed by a thorough discussion, and then more examples for complete clarification. An Instructor's Manual is available.

Name/Title ____________________
College ____________________
City ____________ State ____________ Zip ____________
Text now in use ____________ Enrollment ____________

I understand that I can return any book within ninety days without obligation. Also, if I decide to adopt a book for class use, it will become my complimentary desk copy and the invoice will be cancelled.

Business & Professional Division
ADDISON-WESLEY PUBLISHING COMPANY, INC.
Reading, Massachusetts 01867

Lecture Notes in Economics and Mathematical Systems

New and forthcoming

Volume **100**

Cluster Analysis

A Survey

Edited by *B. S. Duran* and *P. L. Odell*

This volume represents the first book on cluster analysis — the partitioning of a given set into subsets satisfying certain optimality conditions. Similar to factor analysis, it is an important tool for all investigations (e.g. in economics, the life sciences, psychology) which involve the organization of large amounts of empirical data.

1974. vi, 137p. paper/**$7.40**

Volume **103**

Optimal Subset Selection

Multiple Regression, Interdependence, and Optimal Network Algorithms

By *D. E. Boyce, A. Farhi,* and *R. Weischedel*

This book deals with the problem of selecting items (variables, transportation links, facility locations, etc.) from a member set of all available or possible items so as to optimize an objective function subject to a budget or size constraint. Three optimization subset selection problems are discussed: optimal regression analysis; interdependence analysis; optimal network analysis. Computer programs are also included.

1974. xiii, 187p. paper/**$8.20**

Volume **104**

A Neo-Keynesian Theory of Inflation and Economic Growth

By *S. Fujino*

The author investigates processes of cyclical fluctuations, inflation and economic growth, combining short-run analysis with longer term investigation about the economic development. The book develops a new dynamical neoclassical theory of investment and is of particular interest due to today's concern about growth, inflation, and "stagflation."

1974. v, 96p. paper/**$7.40**

Volume **107**

Control Theory Numerical Methods and Computer Systems Modeling

International Symposium, Rocquencourt, June 17-21, 1974

Edited by *A. Bensoussan* and *J. L. Lions*

The contributions to this volume are primarily devoted to the mathematical aspects of control theory and related fields.

1975. approx. 770p. paper/approx. **$21.80**

Complete listing of volumes in
Lecture Notes in Economics and Mathematical Systems 1-100
available from

Springer-Verlag New York Inc.
175 Fifth Avenue
New York, NY 10010

PUBLICATIONS OF THE INTERNATIONAL MONETARY FUND

▼

International Monetary Reform: Documents of the Committee of Twenty. English, French, and Spanish editions. Available on request.

Instruments of Monetary Policy in the United States: The Role of the Federal Reserve System. Ralph A. Young. 1973. US$1.25.

Membership and Nonmembership in the International Monetary Fund: A Study in International Law and Organization. Joseph Gold. 1974. US$10.00.

Voting and Decisions in the International Monetary Fund: An Essay on the Law and Practice of the Fund. Joseph Gold. 1972. US$6.50.

The Stand-By Arrangements of the International Monetary Fund: A Commentary on Their Formal, Legal, and Financial Aspects. Joseph Gold. 1970. US$4.00.

International Reserves: Needs and Availability. 1970. US$6.00.

International Monetary Fund, 1945-1965: Twenty Years of International Monetary Cooperation. (3 vols.) 1969. US$12.50 a set (US$5.00 a volume).

Surveys of African Economies: Vol. 5, Botswana, Lesotho, Swaziland, Burundi, Equatorial Guinea, Rwanda. 1973. All volumes in this series are available in English and French US$5.00 a volume (US$2.50 a volume to university libraries, faculty members, and students).

▼

Address inquiries to the Secretary,
International Monetary Fund,
Washington, D.C. 20431, U.S.A.

Business and Society

BARGAINING WITHOUT BOUNDARIES

The Multinational Corporation and International Labor Relations

Edited by Robert J. Flanagan and Arnold R. Weber

Reports the proceedings of a seminar on the labor problems raised by multinational corporations. Includes reports by representatives of management, labor, and the universities who have had experience with MNCs throughout the world.

Studies in Business and Society series

1974 256 pages Cloth $14.50

NATIONAL MONETARY POLICIES AND THE INTERNATIONAL FINANCIAL SYSTEM

Edited by Robert Z. Aliber

Economists from several nations consider the inadequacies of the Bretton Woods System and focus on the inconsistency between member nations' policies and the international financial system.

Studies in Business and Society series

1974 352 pages Cloth $16.50

LESSONS FROM JAPANESE DEVELOPMENT

An Analytical Economic History

Allen C. Kelley and Jeffrey G. Williamson

Breaking new ground in development studies, this work is a unique economic history of Japan from 1887 to 1915, the "formative years" of Japan's modern economic growth. The authors address the question: Can or should Japan be taken as a "model" for today's developing nations to follow?

1974 320 pages Cloth $15.00

ECONOMICS OF THE FAMILY

Marriage, Children, and Human Capital

Edited by Theodore W. Schultz

Published for the National Bureau of Economic Research

Representing an important recent development in economics, these papers are based on a concept of the family as an entity which makes deliberate economic decisions about the allocation of the scarce resources at its disposal. Originally published as Supplements to the *Journal of Political Economy.*

1974 504 pages Cloth $12.95

The University of Chicago Press

Chicago 60637

emphasis - econ.

ECONOMICS OF SOCIAL ISSUES
Richard H. Leftwich and Ansel M. Sharp, *both of Oklahoma State University*
(paperbound)

MANAGERIAL ECONOMICS: Analysis and Cases, *Third Edition*
W. Warren Haynes, late Dean of Business Administration, *State University of New York at Albany,* and William R. Henry, *Georgia State University*

READINGS IN MANAGERIAL ECONOMICS
Edited by the late W. Warren Haynes, Thomas J. Coyne, *University of Akron,* and Dale K. Osborne, *Southern Methodist University*
(paperbound)

ECONOMICS OF TRANSPORTATION AND LOGISTICS
Marvin L. Fair, *American University,* and Ernest M. Williams, *Columbia University*

THE ESSENTIALS OF MACROECONOMIC ANALYSIS
Richard A. Bilas, *California State College, Bakersfield,* and
Frank J. Alessio, *University of Arizona*

BUSINESS PUBLICATIONS, INC.
13773 North Central Expressway • Suite 1121 • Dallas, Texas 75231

Providing Adequate Retirement Income

Providing Adequate Retirement Income
Pension Reform in the United States and Abroad

by James Schulz, Guy Carrin, Hans Krupp, Manfred Peschke, Elliott Sclar, and J. Van Steenberge

Experts analyze social security pension systems in the United States, Sweden, Germany, Belgium, and Canada, showing how innovative systems abroad suggest alternative options for the United States. They examine the current economic status of the elderly in this country, project future retirement income, and explore alternative policy options, including "dynamic pensions," based upon a nonpoverty definition of income adequacy. *Published for Brandeis University Press.* $15.00

University Press of New England Box 979, Hanover, N. H. 03755

WHAT IS TO BE THE ECONOMICS OF TO-MORROW?

In a new Hobart Paperback

"The Cambridge Revolution: Success or Failure?" Professor Mark Blaug of the University of London argues that:

> "The claim that Cambridge UK is launching a revolution in economic thought, on the scale and in the spirit of the Keynesian Revolution, is largely a piece of misleading advertising, which evaporates as soon as its accomplishments to date are closely examined."

In contrast,

> "Anyone who has been attentive to the resurgence of the neo-classical research programme . . . can hardly doubt that there is life yet in the concepts of maximisation, equilibrium, substitution and all the other tricks of the trade of mainstream economics. Cambridge UK, however it tries, cannot get along without these ideas and in that lies the overwhelming superiority of the neo-classical tradition in economic thought."

An IEA Hobard Paperback **price $6.95**

Other critiques of fashionable thinking from IEA, and especially of the macro-economic monopsony have been

"Macromancy"
(Hobart Paper 55 price $2.50)

in which Douglas Rimmer attacked the misuse of GNP as the measure of economic progress and

"Macro-economic Thinking and the Market Economy"
(Hobart Paper 56 price $2.50)

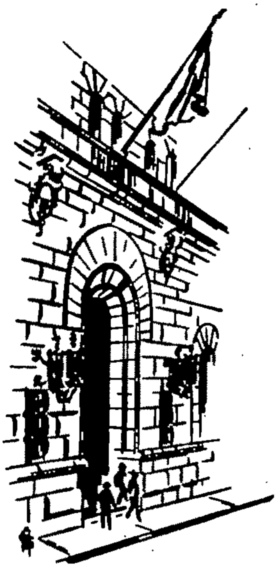
in which Professor L. M. Lachmann rebuked both the latterday Ricardians at Cambridge (UK) and the neo-classicals at Cambridge (USA) for inter alia formalism, which enables some economists "to dispense with individuals, the differences between their minds and the inequality of men in general."

There will be further contributions to the counter-revolution in economy theory from the IEA in 1975. Write for 1975 catalogue

FROM OUR AMERICAN AGENTS:

TRANSATLANTIC ARTS INC., LEVITTOWN, LONG ISLAND, NEW YORK 11756.

INSTITUTE OF ECONOMIC AFFAIRS, 2, LORD NORTH STREET,
LONDON SW1. ENGLAND



MONETARY AGGREGATES AND MONETARY POLICY

A collection of essays examining the role of monetary aggregates in the formulation and execution of monetary policy, written by the Staff of the Federal Reserve Bank of New York.

Single copies are available on request to the Public Information Department of the Federal Reserve Bank of New York, 33 Liberty Street, New York 10045.

JOB OPENINGS FOR ECONOMISTS

A bimonthly listing of job vacancies

Available only to AEA members and to organizations and departments of economics who agree to list their openings.

I would like to subscribe to *Job Openings for Economists* and enclose a check for one year.

1975 Subscription Rates—U.S.A., Canada, and Mexico (first class): $6.00 junior member, $12.00 others; foreign airmail: $12.00 junior member, $18.00 others.

Name ______________________________

Address ______________________________

Check one:

☐ I am a member of the American Economic Association.
☐ I am not a member but would like to become one. My application is enclosed.
☐ (For organizations) We agree to list our vacancies in JOE.

THE AMERICAN ECONOMIC ASSOCIATION
1313 21st Avenue South
Nashville, Tennessee 37212

An Opportunity For Departure

Yes, an opportunity to pursue a bold, new, and effective approach to economics. A refreshing departure from past approaches to economic concepts—approaches that failed miserably in their basic purpose—that of teaching economics to students.

ECONOMICS—The Science of Common Sense by E. V. Bowden is designed to accomplish one basic objective—to present economic principles as simply and as clearly as possible. To this end, we anticipate the greatest success. You'll agree when you examine ECONOMICS—The Science of Common Sense.

SOUTH-WESTERN PUBLISHING CO.

Cincinnati, OH 45227 • Dallas, TX 75229 • Palo Alto, CA 93404
West Chicago, IL 60185 • Pelham Manor, NY 10803 • Brighton, England

Lexington Books

D. C. HEATH AND COMPANY
125 Spring Street
Lexington, Massachusetts 02173

HEATH

CASH MANAGEMENT
An Inventory Control Limit Approach
Richard B. Homonoff and David W. Mullins, Jr.
128 pp. $12.00 1975

AN ECONOMETRIC MODEL OF THE LIFE INSURANCE SECTOR OF THE U.S. ECONOMY
J. David Cummins, *University of Pennsylvania*
272 pp. $19.50 1974

FINANCIAL RATIOS AND INVESTMENT RESULTS
Donald M. Peterson
160 pp. $16.00 1974

TOWARD A NEW WORLD TRADE POLICY: THE MAIDENHEAD PAPERS
C. Fred Bergsten, *The Brookings Institution*
In Press Spring 1975

A GUIDE TO INTERNATIONAL MONETARY REFORM
George N. Halm
In Press Spring 1975

INVESTMENT PORTFOLIO DECISION-MAKING
edited by James L. Bicksler, *Rutgers University;* Paul A. Samuelson, *MIT*
384 pp. $15.00 1974

OPTIMIZATION OF BANK PORTFOLIOS
William F. Beazer, *University of Virginia*
In Press Winter/Spring 1975

REGULATION INDUCED DISTORTIONS
Charles W. Needy, *University of Houston*
144 pp. $12.50 1975

THE THEORY OF TRANSPORTATION ECONOMICS
James T. Kneafsey, *MIT*
In Press Spring 1975

1973-1985 IN FIGURES
Supplement to 1985: Interindustry Forecasts of The American Economy
Clopper Almon, Jr., Margaret R. Buckler, Lawrence M. Horwitz, Thomas C. Reimbold, *University of Maryland*
208 pp. $16.00 1974

TELEVISION ECONOMICS
Bruce Owen, *Stanford University;* Jack Beebe, *FRS Associates;* Willard Manning, *Harvard University*
240 pp. $15.00 1974

INTERNATIONAL COMMODITY AGREEMENTS
Setting, Performance, and Prospects
Alton D. Law, *Western Maryland College*
144 pp. $12.50 1975

THE ENVIRONMENTAL CRISIS AND CORPORATE DEBT POLICY
Marvin E. Ray, *Valdosta State College*
123 pp. $11.00 1974

IASP titles in economics

John Kenneth Galbraith and the Lower Economics

Second, Expanded Edition

Myron E. Sharpe, **Editor of CHALLENGE: The Magazine of Economic Affairs**

"The best review I have yet seen of Galbraith's works . . . Sharpe [in the first edition] has succeeded in highlighting Galbraith's contribution to economic thought, while exposing to critical scrutiny the weakest parts of his analysis. I'm happy to recommend this book.

—**Robert L. Heilbroner**

This edition contains a new section evaluating Galbraith's **Economics and the Public Purpose**, and includes a lengthy interview with Galbraith.

$7.95

The Socialist Economies of the Soviet Union and Europe

Marie Lavigne, **University of Paris**

"Marie Lavigne's book is both thoroughly competent and of a sufficiently high standard to raise French Sovietology to the level of English or American works on the same subject." —**Revue Economique**
(Reprinted in **Economic Journal**)

This encyclopedic study devotes major attention to the Soviet economic system, and is unusually detailed in describing the other socialist economies and in drawing upon them for important cross-national comparisons.

Dr. Lavigne's publications include **Capital in the Soviet Economy, The Problem of Prices in the Soviet Union** (with Prof. H. Denis), and **COMECON: From Cooperation to Integration of the Socialist Economies.**

$20.00

Monetary Cooperation Between East and West

Adam Zwass, **Vienna Institute for Comparative Economic Studies**

Foreword by George Garvy

A systematic study of the instruments and processes of East-West financial relations based on a critical analysis of the money, credit and foreign trade systems of the East European countries and of their Council for Mutual Economic Assistance.

Dr. Zwass occupied managerial positions in the central banking systems of Poland and the USSR. From 1963 to 1968 he was a Counsellor in the COMECON Secretariat in Moscow, where he was responsible for the work of the International Bank for Economic Cooperation.

$15.00

iasp
INTERNATIONAL ARTS & SCIENCES PRESS Inc.
901 NORTH BROADWAY
WHITE PLAINS, NEW YORK 10603

Essays in Honor of Jan Tinbergen

Edited by Willy Sellekaerts, **The Wharton School, University of Pennsylvania**

Three volumes of original articles by some of the world's leading economists. Each volume includes a selected bibliography of Tinbergen's work, as well as an introduction which appraises Tinbergen's major contributions to economics. The three volumes are: **International Trade and Finance; Economic Development and Planning; Econometrics and Economic Theory.**

$20.00 per volume

Self-Governing Socialism: A Reader

Edited by Branko Horvat, Mihailo Markovic and Rudi Supek

Three distinguished Yugoslav scholars have selected seminal articles, book extracts, and documents from 20 countries in order to trace the theory and practice of self-governing socialism from the early nineteenth century to the present. Much of the material is being made available in English for the first time. Volume I: Historical Development; Social and Political Theory. Volume II: Sociology and Politics; Economics.

Vol. I	(cloth)	$21.50	Vol. II	(cloth)	$18.50
	(paper)	$9.50		(paper)	$7.95

Advertising and Socialism

The Nature and Extent of Consumer Advertising in the Soviet Union, Poland, Hungary and Yugoslavia

Philip Hanson, **University of Birmingham**

The most thorough and detailed study to date of consumer advertising in socialist economies. The author employs a substantial amount of new data — much of which has been obtained from extensive interviewing of managerial personnel — to examine and compare the organization, nature, and purposes of advertising in four socialist countries.

Dr. Hanson is the author of **The Consumer in the Soviet Economy.**

$15.00

Land Reform and Economic Development in China

A Study of Institutional Change and Development Finance

Victor D. Lippit, **University of California at Riverside**

A new analysis that makes a significant contribution to the economics of the Chinese Revolution and to development theory generally.

Professor Lippit shows how a poor country managed simultaneously to raise the share of national income devoted to investment and increase living standards. He focuses on the use of the "surplus" — the capital or property share of national income — to redirect income flows, thereby to shift consumption from luxuries to necessities and increase public investment.

$15.00

Still a Dream
The Changing Status of
Blacks Since 1960
Sar A. Levitan,
William B. Johnston,
and Robert Taggart

The most comprehensive review of the changing social and economic status of Blacks in the United States during the critical years since 1960, this definitive work investigates how changes occurred, how much change took place, and how public policy affected the transformations. The authors argue that conditions of the 1970s have offset some of the gains of the previous decade. $15.00

Harvard University Press
79 Garden St., Cambridge,
MA 02138

Introduction to Urban Economics

DOUGLAS M. BROWN, *Georgetown University*

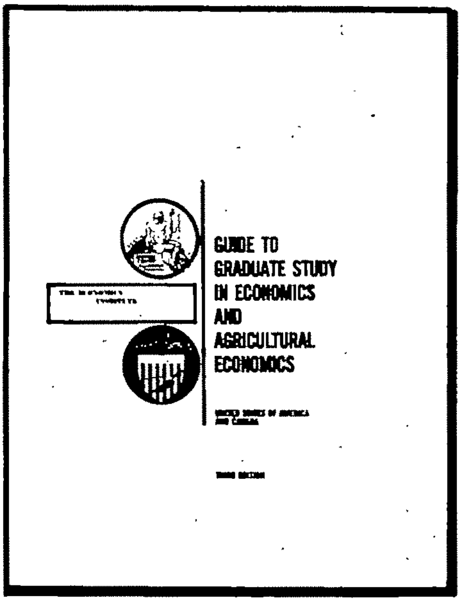
This textbook provides the student with a knowledge of, and the tools for, analysis of contemporary urban economic problems. The history, growth, and development of cities is examined in detail, with particular emphasis on the relationship between urban problems and economic analysis.

- Provides complete coverage of interurban and intraurban economic analysis, requiring only a background in economic principles.
- Surveys the economic history of cities from tribal organizations to the American city today, with special attention given to colonial cities.
- Integrates theoretical and empirical analysis of urban markets with a lively discussion of contemporary problems.
- Examines the causes and possible solutions for critical urban problems from the point of view of economic analysis.
- Presents a wide spectrum of theoretical and empirical analysis with a minimal use of mathematics.

301 pages, $9.95

ACADEMIC PRESS
A Subsidiary of Harcourt Brace Jovanovich, Publishers
111 Fifth Avenue, New York, N.Y. 10003
24-28 Oval Road, London NW1 7DX

Announcing...



GUIDE TO GRADUATE STUDY IN ECONOMICS AND AGRICULTURAL ECONOMICS

Guide to Graduate Study in Economics and Agricultural Economics

in the United States
of America
and Canada
THIRD EDITION

Designed to provide students anticipating graduate study in economics and agricultural economics, and their advisors, with information on available graduate training programs.

Includes descriptions of 236 graduate programs, supplemented by comparative data and information for prospective students, domestic and foreign.

Published by the Economics Institute under the auspices of the American Economic Association, and the American Agricultural Economics Association.

Distributed by:
RICHARD D. IRWIN INC.,
1818 Ridge Road,
Homewood, Illinois 60430

PRICE: $7.50 per copy

TECHNICAL CHOICE, INNOVATION AND ECONOMIC GROWTH

Essays on American and British Experience in the Nineteenth Century

PAUL ALLAN DAVID, *Professor of Economics, Stanford University*

Self-contained, related essays focusing on specific industries and individual productive units—from a New England textile mill to a mid-Victorian grain farm. By "one of the leaders of the movement which has been dubbed 'The New Economic History.' His writings invariably display a mastery of historical literature relevant to the problem at hand."—NATHAN ROSENBERG, *The University of Wisconsin*
Cloth $21.00 Paper $5.95

LATE NINETEENTH-CENTURY AMERICAN DEVELOPMENT

A General Equilibrium History

JEFFREY G. WILLIAMSON, *Professor of Economics, University of Wisconsin*

Professor Williamson probes issues central to American economic history, such as the role of the railroads, the distribution of income, development of the frontier and capital market development. His findings often differ from the established "truths" because here for the first time this critical period is analyzed with sophisticated models and extensive data. $19.50

RATIONAL ECONOMIC MAN

A Philosophical Critique of Neo-Classical Economics

MARTIN HOLLIS, *Lecturer in Philosophy, University of East Anglia*

EDWARD J. NELL, *Professor of Economics, New School for Social Research*

The authors attack the neo-classical theories of economics and the positivist theories of knowledge upon which they rely. They argue convincingly for a rationalist philosophy and a classical or Marxian form of economics. $14.95

Now in paperback—two major contributions to economic literature by M. H. DOBB

WELFARE ECONOMICS AND THE ECONOMICS OF SOCIALISM $5.95

THEORIES OF VALUE AND DISTRIBUTION SINCE ADAM SMITH $5.95

Recently published

INTERNATIONAL BUSINESS AND CZECHOSLOVAKIA 1918-1938
A. TEICHOVA $27.50

PRODUCTION CONDITIONS IN INDIAN AGRICULTURE
K. BHARADWAJ Cloth $12.50 Paper $5.95

EXPORTING WORKERS: THE TURKISH CASE
S. PAINE Cloth $13.95 Paper $6.95

Cambridge University Press
32 East 57th Street, New York, N.Y. 10022

The American Economic Review

PAPERS AND PROCEEDINGS

OF THE

Eighty-seventh Annual Meeting

OF THE

AMERICAN ECONOMIC ASSOCIATION

San Francisco, California, December 28–30, 1974

MAY 1975

AMERICAN ECONOMIC ASSOCIATION

Organized at Saratoga, New York, September 9, 1885

PAST OFFICERS

Presidents

Francis A. Walker,* M.I.T., 1886–92
Charles F. Dunbar,* Harvard, 1893
John B. Clark,* Columbia, 1894–95
Henry C. Adams,* Michigan, 1896–97
Arthur T. Hadley,* Yale, 1898–99
Richard T. Ely,* Wisconsin, 1900–01
Edwin R. A. Seligman,* Columbia, 1902–03
Frank W. Taussig,* Harvard, 1904–05
Jeremiah W. Jenks,* Cornell, 1906–07
Simon N. Patten,* Pennsylvania, 1908
Davis R. Dewey,* M.I.T., 1909
Edmund J. James,* Illinois, 1910
Henry W. Farnam,* Yale, 1911
Frank A. Fetter,* Princeton, 1912
David Kinley,* Illinois, 1913
John H. Gray,* Minnesota, 1914
Walter F. Willcox,* Cornell, 1915
Thomas N. Carver,* Harvard, 1916
John R. Commons,* Wisconsin, 1917
Irving Fisher,* Yale, 1918
Henry B. Gardner,* Brown, 1919
Herbert J. Davenport,* Cornell, 1920
Jacob H. Hollander,* Johns Hopkins, 1921
Henry R. Seager,* Columbia, 1922
Carl C. Plehn,* California, 1923
Wesley C. Mitchell,* Columbia, 1924
Allyn A. Young,* Harvard, 1925
Edwin W. Kemmerer,* Princeton, 1926
Thomas S. Adams,* Yale, 1927
Fred M. Taylor,* Michigan, 1928
Edwin F. Gay,* Harvard, 1929
Matthew B. Hammond,* Ohio State, 1930
Ernest L. Bogart,* Illinois, 1931
George E. Barnett,* Johns Hopkins, 1932
William Z. Ripley,* Harvard, 1933
Harry A. Millis,* Chicago, 1934
John M. Clark,* Columbia, 1935
Alvin S. Johnson,* New School, 1936
Oliver M. W. Sprague,* Harvard, 1937
Alvin H. Hansen, Harvard, 1938
Jacob Viner,* Chicago, 1939
Frederick C. Mills,* Columbia, 1940
Sumner H. Slichter,* Harvard, 1941
Edwin G. Nourse,* Brookings, 1942
Albert B. Wolfe,* Ohio State, 1943
Joseph S. Davis, Stanford, 1944
I. L. Sharfman,* Michigan, 1945
E. A. Goldenweiser,* Institute for Advanced Study, 1946
Paul H. Douglas, Chicago, 1947
Joseph A. Schumpeter,* Harvard, 1948
Howard S. Ellis, California, 1949
Frank H. Knight,* Chicago, 1950
John H. Williams, Harvard, 1951
Harold A. Innis,* Toronto, 1952
Calvin B. Hoover, Duke, 1953
Simon Kuznets, Pennsylvania, 1954
John D. Black,* Harvard, 1955
Edwin E. Witte,* Wisconsin, 1956
Morris A. Copeland, Cornell, 1957
George W. Stocking, Vanderbilt, 1958
Arthur F. Burns, Columbia, 1959
Theodore W. Schultz, Chicago, 1960
Paul A. Samuelson, M.I.T., 1961
Edward S. Mason, Harvard, 1962
Gottfried Haberler, Harvard, 1963
George J. Stigler, Chicago, 1964
Joseph J. Spengler, Duke, 1965
Fritz Machlup, Princeton, 1966
Milton Friedman, Chicago, 1967
Kenneth E. Boulding, Colorado, 1968
William J. Fellner, Yale, 1969
Wassily Leontief, Harvard, 1970
James Tobin, Yale, 1971
John Kenneth Galbraith, Harvard, 1972
Kenneth J. Arrow, Harvard, 1973
Walter W. Heller, Minnesota, 1974

Secretaries

Richard T. Ely,* 1886–92
Edward A. Ross,* 1893
Jeremiah W. Jenks,* 1894–96
Walter F. Willcox,* 1897–99

Treasurers

Edwin R. A. Seligman,* 1886–90
Frederick B. Hawley,* 1891–95
Charles H. Hull,* 1896–99

Secretary-Treasurers

Charles H. Hull,* 1900
Frank A. Fetter,* 1901–06
Winthrop M. Daniels,* 1907–08
Thomas N. Carver,* 1909–13
Allyn A. Young,* 1914–20
Ray B. Westerfield,* 1921–25
Frederick S. Deibler,* 1925–35
James Washington Bell,* 1936–61
Harold F. Williamson, 1962–70

Editors

Davis R. Dewey,* 1911–40
Paul T. Homan,* 1941–51
Bernard F. Haley, 1952–62
John G. Gurley, 1962–68

*Deceased

OFFICERS FOR 1975

President

Robert Aaron Gordon,
University of California, Berkeley

President-elect

Franco Modigliani,
Massachusetts Institute of Technology

Vice-Presidents

Martin Bronfenbrenner, Duke University
Charles L. Schultze, University of Maryland

Secretary-Treasurer and Editor of Proceedings

Rendigs Fels, Vanderbilt University

Managing Editor of the American Economic Review

George H. Borts, Brown University

Managing Editor of the Journal of Economic Literature

Mark Perlman, University of Pittsburgh

Elected Members of the Executive Committee

Irma G. Adelman, University of Maryland
Robert L. Heilbroner, New School for Social Research
Andrew F. Brimmer, Harvard University
Paul M. Sweezy, Monthly Review
Carolyn Shaw Bell, Wellesley College
Burton A. Weisbrod, University of Wisconsin

Ex Officio Members

Kenneth J. Arrow, Harvard University
Walter W. Heller, University of Minnesota

Counsel

Donald F. Turner, Harvard University

THE AMERICAN
ECONOMIC REVIEW

VOL. LXV MAY, 1975 NUMBER 2

PAPERS AND PROCEEDINGS

OF THE

Eighty-seventh Annual Meeting

OF THE

AMERICAN ECONOMIC ASSOCIATION

San Francisco, California
December 28–30, 1974

Program Arranged by ROBERT AARON GORDON
Papers and Proceedings Edited by RENDIGS FELS
and C. ELTON HINSHAW

Copyright American Economic Association, 1975

PRINTED BY GEORGE BANTA COMPANY, INC.

Publication Office: Menasha, Wisconsin 54952

Executive Office: American Economic Association
1313 21st Avenue South, Nashville, Tennessee 37212

Inquiries and other communications regarding membership, meetings, and the general affairs of the Association, as well as orders for the publications and permission to reprint, should be addressed to Rendigs Fels, Secretary of the American Economic Association, 1313 21st Avenue South, Nashville, Tennessee 37212.

Changes of address should also be sent to the Secretary at the address given above. To be effective, notice must reach the office by the 1st of the month prior to the month of publication. The Association cannot be responsible for copies lost due to failure to report change of address in time.

Second-class postage paid at Nashville, Tennessee and at additional mailing offices. Printed in U.S.A.

No responsibility for the views expressed by authors in this *Review* is assumed by the editors or the publishers, the American Economic Association.

Membership dues of the Association are $23 a year, $18.40 of each year's dues being in payment of a year's subscription to both the *American Economic Review* and the *Journal of Economic Literature.* The *American Economic Review* is published five times a year (March, May, June, September, and December) with four quarterly issues, the *Papers and Proceedings* of the annual meetings, the *Directory*, and supplements, and is sent to all members of the American Economic Association as one of the privileges of membership, as is the *Journal of Economic Literature* (four quarterly issues per year). Subscriptions by nonmembers (libraries, institutions, etc.) are $34.50 annually. Only subscriptions to both publications will be accepted. Single copies of the *Review* and *Journal* may be ordered from the Executive Office.

Price $9.50

CONTENTS

Editors' Introduction.......................... *Rendigs Fels and C. Elton Hinshaw* vii

PAPERS

Richard T. Ely Lecture

Income Distribution—Can Economists Help?.................... *Alice M. Rivlin* 1

The Economics of Labor: An Assessment of Recent Research

The Challenge of Dual and Radical Theories of the Labor Market to Orthodox Theory.. *Glen G. Cain* 16
Economic Analysis of Trade Unionism....................... *George E. Johnson* 23
Economics of Time and Labor Supply............................ *H. G. Lewis* 29
Discussion... *Paula E. Stephan* 35

Past Developments and Future Prospects for Ethnic Minority Groups

Three Phases of Cliometric Research on Slavery and its Aftermath... *Robert William Fogel* 37
Public Policy and Black Economic Progress: A Review of the Evidence.. *Bernard E. Anderson and Phyllis A. Wallace* 47
Factors Affecting the Future Economic Prospects of Minorities... *David H. Swinton* 53
Discussion... *Ronald L. Trosper* 59
.... *Daniel Fusfeld* 60

The Human Capital Approach: An Appraisal

Human Capital Theory: Education, Discrimination, and Life Cycles... *Finis Welch* 63
The Problem with Human Capital Theory—A Marxian Critique.. *Samuel Bowles and Herbert Gintis* 74

Supply and Mobility of Women Economists

Job Search and Affirmative Action............. *Alice H. Amsden and Collette Moser* 83
Women Economists: Career Aspirations, Education, and Training.. *Myra H. Strober* 92
Two Supply Curves for Economists? Implications of Mobility and Career Attachment of Women.................................... *Barbara B. Reagan* 100

Joint Luncheon Address

Inflation and the International Monetary Situation......... *H. Johannes Witteveen* 108

Understanding World Inflation

Monetarist and Keynesian Models of the Transmission of Inflation.. *William H. Branson* 115
Understanding World Inflation.................................. *George L. Perry* 120
Monetary Growth and the Long-Run Rate of Inflation.......... *Richard T. Selden* 125
Interpreting Recent World Inflation.......................... *Ronald L. Teigen* 129

Domestic Adjustment to Balance of Payments Disequilibrium

The Payments Adjustment Process and the Exchange Rate Regime: What Have We Learned?.................................... *Marina v.N. Whitman* 133
Discussion... *Rudiger Dornbusch* 147
.... *William J. Fellner* 148
.... *Alexandre Kafka* 151
.... *Paul A. Volcker* 153

25 Years After the Rediscovery of Money: What Have We Learned?

Recent Developments in Monetary Theory...................... *Stanley Fischer* 157

Empirical Monetary Macroeconomics: What Have We Learned in the Last 25 Years? .... *W. C. Brainard and R. N. Cooper* 167
Discussion........ *Milton Friedman* 176
.... *Franco Modigliani* 179

A Critical Look at the Keynesian Model: Theory and Application

The Coordination of Economic Activities: A Keynesian Perspective........ *Robert Clower and Axel Leijonhufvud* 182
The Keynesian Revolution Reconsidered........ *Robert Eisner* 189
Keynesian Models of Recession and Depression........ *James Tobin* 195

Personal Savings in the Postwar World: Implications for The Theory of Household Behavior

Towards a Theory of Saving Behavior........ *F. Thomas Juster and Lester D. Taylor* 203
Saving Behavior in Western Germany and the United States........ *Burkhard Strumpel* 210
Postwar U.S. Consumption, Consumer Expenditures, and Saving..*Michael R. Darby* 217
Discussion........ *Thomas Mayer* 223

Microeconomic Theory: Conflict and Contract

A Contractarian Paradigm for Applying Economic Theory........ *James M. Buchanan* 225
Problems vs. Conflicts: Economic Theory and Ideology........ *Duncan K. Foley* 231
Law and Economic Theory: An Economist's View........ *Alvin K. Klevorick* 237

Sources and Uses of Panels of Microdata

The National Longitudinal Surveys: New Vistas for Labor Market Research........ *Herbert S. Parnes* 244
Using Survey Data from the University of Michigan's Survey Research Center........ *James N. Morgan* 250
Potential Applications of Census Bureau Economic Series in Microdata Analysis........ *Shirley Kallek* 257
The Experimental Panel Data Resources at the Poverty Institute Data Center........ *Katherine Dickinson and Harold W. Watts* 263

The Behavior of Firms in Oligopolistic Markets: In Theory and Practice

Firm Decision-making Processes and Oligopoly Theory........ *Paul L. Joskow* 270
Oligopoly Theory, Communication, and Information........ *Martin Shubik* 280

Emerging Structural Changes in Transport and Public Utilities

Northeast Railroads: Restructured or Nationalized?........ *Alexander Lyall Morton* 284
Transportation Controls and the Spatial Structure of Urban Areas........ *Gerald S. Goldstein and Leon N. Moses* 289
Regulation and the Financial Condition of the Electric Power Companies in the 1970's........ *Paul L. Joskow and Paul W. MacAvoy* 295

A Reassessment of Development Economics

Development Economics—A Reassessment of Goals........ *Irma Adelman* 302
The Structuralist Approach to Development Policy........ *Hollis B. Chenery* 310
State of Development Economics........ *Stephen A. Resnick* 317

What Have We Learned from Economic Growth Theory?

Implications of Growth Theory for Trade and Development........ *Ronald Findlay* 323
The Theory of Economic Growth: A Critique and Reformulation..*Donald J. Harris* 329
Growth Theory from an Evolutionary Perspective: The Differential Productivity Puzzle........ *Richard R. Nelson and Sidney G. Winter* 338

China and India: Development During the Last 25 Years
Chinese and Indian Development: An Interdisciplinary Environmental Analysis... *Barry Richman* 345
China and India: Contrasting Experiences in Economic Development... *Thomas E. Weisskopf* 356
Discussion... *Padma Desai* 365
... *John G. Gurley* 368

East-West Technological Transfer Issues and Problems
International Technology Transfer: Forms, Resource Requirements, and Policies... *Edwin Mansfield* 372
The Economics of East European Technology Imports from the West... *Edward A. Hewett* 377
Problems of Technology Absorption in Chinese Industry... *Thomas G. Rawski* 383

Taxation of Natural Resources
Incidence of World Taxes on Natural Resources with Special Reference to Bauxite... *Malcolm Gillis and Charles E. McLure, Jr.* 389
U.S. Taxes on Energy Resources... *Gerard M. Brannon* 397
Discussion... *Arthur W. Wright* 405

Optimizing Models of Public Decision Making
Purposive Models of Legislative Behavior... *John A. Ferejohn and Morris P. Fiorina* 407
An Evolutionary and Institutional View of the Behavior of Public and Private Companies... *Marc J. Roberts* 415

The Principles Course: What Should Be in It and Where Should It Be Going?
Some Observations on the Learning of Economics... *Kenneth E. Boulding* 428
Some Comments on the Principles Course... *John G. Gurley* 431
New Approaches to Teaching the Principles Course... *W. Lee Hansen* 434

PROCEEDINGS

Minutes of the Annual Meeting... 441

Minutes of the Executive Committee Meetings... 448

Reports
Secretary... *Rendigs Fels* 454
Treasurer... *Rendigs Fels* 460
Finance Committee... *Beryl W. Sprinkel* 464
Auditors... *Arthur Andersen & Co.* 467
Managing Editor, American Economic Review... *George H. Borts* 475
Managing Editor, Journal of Economic Literature... *Mark Perlman* 481
Director, Job Openings for Economists... *C. Elton Hinshaw* 488
Committee on the Status of Women in the Economics Profession... *Barbara B. Reagan* 490
Summer Program in Economics for Minority Students... *Marcus Alexis* 502
Committee on Economic Education... *G. L. Bach* 506
Economics Institute Policy and Advisory Board... *Edwin S. Mills* 508
Representative to the International Economic Association... *Fritz Machlup* 509
Representative to the National Bureau of Economic Research... *Willard L. Thorp* 511
Census Advisory Committee... *Lee E. Preston* 513

THE purpose of the American Economic Association, according to its charter, is the encouragement of economic research, the issue of publications on economic subjects, and the encouragement of perfect freedom of economic discussion. The Association as such takes no partisan attitude, nor does it commit its members to any position on practical economic questions. It is the organ of no party, sect, or institution. Persons of all shades of economic opinion are found among its members, and widely different issues are given a hearing in its annual meetings and through its publications. The Association, therefore, assumes no responsibility for the opinions expressed by those who participate in its meetings. Moreover, the papers presented are the personal opinions of the authors and do not commit the organizations or institutions with which they are associated.

Editors' Introduction

THIS volume contains the papers and proceedings of the eighty-seventh annual meeting of the American Economic Association. "Proceedings" concern the business activities of the Association—the annual membership meeting, the twice yearly meetings of the Executive Committee, and reports of various Association officers and committees—and, as with the "Notes" section in each issue of the *American Economic Review*, are published in the hope that members will be informed of, and will participate in, the Association's affairs.

The "Papers," which constitute the greater part of this volume, are roughly equivalent to two regular issues of the *American Economic Review*, but are published under quite different procedures. An explanation of current Association policies concerning papers for the annual meetings may therefore be of some interest.

Early in the year the Association's President-elect (in 1974 R. A. Gordon, in 1975 Franco Modigliani) sets in motion the process which culminates in the December presentation and May publication of the papers. After consultation and comments, both volunteered and solicited, from a wide range of persons, the President-elect decides on the topics of sessions (roughly twenty) at which papers will be presented, sets limits on the length of papers and invites persons to organize these sessions. Each session organizer in turn invites, again after consultation, several persons (usually two or three in number) to give papers on the topic of the session in question, and asks others to give comments on the papers. The papers, or rough drafts, are supposed to be sent to the editors of the *Papers and Proceedings* by November 15; we then read them and return them to their authors with comments and suggestions. In late December the papers and comments are read at the Association's meetings; we ask the authors to send us the final versions of their papers by January 2. After giving the final versions further scrutiny, we then in almost all cases send them to the printer to be published. (We do not generally publish discussants' comments, however, unless there is some special reason to do so.)

Thus, the rules under which these papers are published are quite different from those concerning articles appearing in regular issues of the *Review*. The papers published here are invited, not contributed; except in unusual circumstances they must be less than 4,000 (and sometimes 3,000 or even 2,000) words in length; each author is asked to avoid highly technical discussion and instead to contribute a paper which will be intelligible to economists not specializing in the subject it discusses. While we do edit (sometimes extensively) most papers to improve content and style and to satisfy space constraints, we do not subject papers to any refereeing process, and publication of any paper which we receive prior to the final printer's deadline and which satisfies (or can be cut to satisfy) space requirements is virtually guaranteed.[1]

These policies are advantageous in a number of respects. The volume of papers can be published without the long delays inherent in publication by regular journals, and it consists of relatively short papers which concern a wide variety of subjects and can be understood by nonspecialists. Authors get a chance to report briefly on research just completed, to discuss topical subjects in an informal and engaging way, and to summarize longer forthcoming publications; readers get a chance to browse among a large number of articles which are outside their major areas of interest but which are not as specialized or ponderous as those sometimes found in regular journals. And while the papers are not scrutinized in the way regular journal articles are, they are (or are at least intended to be!) nontechnical, and thus do not need the thorough refereeing usual for articles submitted to journals.

RENDIGS FELS
C. ELTON HINSHAW

---

[1] We would refuse to publish such a paper only if we felt that it was utterly without merit; no paper has yet been rejected on these grounds. However, the Executive Committee has established another ground for rejection: if a paper cannot be cut to meet space requirements, we ask the author either to authorize us to submit it to the managing editor of the *Review* to be considered for publication (subject to the usual refereeing process) in a regular issue of that journal, or to withdraw the paper and submit it elsewhere.

RICHARD T. ELY LECTURE

Income Distribution—Can Economists Help?

By ALICE M. RIVLIN*

Since it is always safer to make predictions in a field other than one's own, I want to start this lecture with a political prediction, namely, that income shares and distributional aspects of public actions are going to become a major focus of policy debate in the next few years. I am not predicting how that debate will come out—just that it will take place. I want to talk this evening about how economists can be useful in raising the debate from a shouting match to something approaching an informed argument over real issues.

In choosing this topic I am implicitly contending that economists ought to try to make themselves useful and that public policy will be "better" if they do. Both these points are arguable, but I do not intend to get into that argument. Nor do I feel it necessary to defend the proposition that income is a reasonable proxy, if not for happiness, at least for economic well-being, and hence its distribution is worth talking about.

There are several reasons for thinking that the public and its political representatives are likely to be increasingly concerned about income shares. First is the continuing fact that instant communication, especially televisual communication, has made everyone more conscious of what others have and do not have, while at the same time rising real incomes and constant relative shares have widened the absolute gap between the fairly rich and the fairly poor. It is not just those at the bottom of the heap who are more conscious than ever of what they are missing. Increasingly—perhaps by reaction to the political emphasis on extreme poverty in the 1960's—it is those near the middle who seem most militantly concerned about their share. The double devastation of inflation and recession, moreover, exacerbates the concern with income shares, since everyone, including stock brokers, feels abused when real income is falling.

The second reason for the mounting concern over income shares is the apparent rise in group identities of Americans and the new political sophistication of these groups in pressing their advantage through the law, collective bargaining, public relations, political influence, and even civil disobedience. Groups learn effective tactics from each other and we can look forward in the next few years to organized group pressure at all levels by blacks, Chicanos, native Americans, women, Polish-Americans, old people, young people, blue-collar workers, public employees—and maybe even middle-aged white males.

Third, there are some demographic-social forces that may move the Lorenz curve of family income farther from the 45-degree line and increase public awareness and political concern with inequality. The rising average age of family heads could have this effect, but a more dramatic disequalizing force may well be the changing status of women. The proportion of families headed by women has been increasing and is likely to continue to do so. This need not be seen as a pathological development—it may reflect an increasing capacity of women to fend for themselves in the marketplace and a real increase in the welfare of those involved, but it clearly makes the distribution of family income more unequal than it would otherwise be. A more important effect of women's changing status may be to increase the inequality of income among husband-wife families. As long as wives have been secondary earners in every sense of the word, their earnings have had a generally equalizing effect on family income. But this effect would be reversed if a major increase occurred in the proportion of wives with a permanent, lifelong attachment to the labor force and an earnings distribution similar to that of husbands. People with high earning capacity tend to marry each other, and this tendency might well be increased if job status became a more central part of the average woman's life. In a world in which professionals and managers married each other, while unskilled workers did the same, and in which male and female patterns of working and earnings were similar, the concentration of money income among husband-wife families at the high end of the scale would be even greater than it is now, and the income share of those at the bottom would be lower.

---

* Member of the Economics Studies Program of The Brookings Institution. The views expressed are mine and not necessarily those of the officers, trustees, or other staff members of The Brookings Institution.

The political focus on distributional questions that I expect for the next few years need not be an overt fight about redistributing money income as such through taxes or transfers or other policies explicitly designed to do just that. It might take the form of increased concern with and friction over the distributional impact of other policies sure to be on the national agenda, such as health insurance, energy conservation, pollution, and measures to combat inflation and recession.

Assuming that I am right in this political speculation, it seems to me that there are four sets of questions that the public and the politicians might reasonably ask a useful group of economists.

1. What is happening?
   How great is income inequality and which way is it moving?
   What are the effects of current government programs?
2. What would be better?
   Presumably no one wants economists telling them what the distribution of income ought to be, but they might reasonably expect them to be helpful in defining alternative norms and discussing how varying degrees of equality might be expected to affect other objectives.
3. Why are things as they are?
   Economists worth their pay ought to provide a workable theory that explains why the distribution of income has the shape it has and why it is changing or not changing.
4. What can we do about it?
   Useful economists ought to spell out some policy options and estimate their effects, not only on the distribution of income but on the size of the pot to be distributed.

Clearly this is a large order and the questions are intimately interrelated. One needs a theoretical framework to explain facts and to make predictions either about the effect of alternative distributions on economic behavior or about the efficacy of particular policies for redistributing income. Nevertheless, it may be useful to address these questions sequentially and ask what economists now have to say about each of them and where we might profitably concentrate our future efforts.

I. What is Happening?

To the layman the question may sound simple: Is the distribution of income becoming more equal or more unequal? But one of the main contributions of economists has been to point out that the question is anything but simple and that it cannot be answered without a painstaking effort to define some terms.

First, it matters what *units* one is talking about. The distribution of income among individuals is meaningless because people live in groups and pool their incomes, so we must look at the income of families, households, or perhaps "spending units." These groupings, however, are of varying size and composition. Hence, in considering the degree of inequality some adjustment ought to be made—although it usually is not—for the number of people in the unit and perhaps for the varying needs of different kinds of people. Second, it makes a difference what is *counted as income*. It matters whether one measures income before or after taxes and whether one sticks to cash or adds estimates of income in kind and services of major durable goods such as houses. In particular, it matters how one counts the services of government. It seems absurd to count cash transfers, such as social security, as income while omitting near-cash benefits, such as food stamps. But should one also count the use of roads and the putative protection of nuclear missiles? Third, it matters what *time* period one is talking about. A year is an arbitrary accounting period that is too short for some purposes and too long for others. Short-period inequality is no less "real" than long-period inequality—a kid without a dime cannot get a stick of bubble gum with his permanent income—but it may cause a different degree of concern and call for different remedies. Finally, there are many *measures of inequality* and they may move in opposite directions at the same time. The one most people seem to find comprehensible, the Lorenz curve, can give ambiguous results when the real world is ambiguous. It can show, for example, that the size distribution of income has moved farther from equality at the bottom and closer to it at the top. Moreover, a lot of people who say they are concerned about inequality are really concerned about poverty or income inadequacy, not about inequality per se, and poverty has to be defined somehow, either as inability to afford a specified collection of goods and services at current prices or as a state of deprivation relative to what some major fraction of the population enjoys.

Economists have a smashing record of pointing out all of these problems in measuring income inequality, but, with one major exception, a rather dismal success rate in collecting the refined data we all keep saying is so necessary to answer the questions. The exception is the quantum leap in knowledge about the economic status of the poor that has taken place in the last decade, in large part through the tireless and often unglamorous efforts of economists and other social scientists to refine definitions and goad the government into collecting new information about the low-income population—who they are, where they are, and how they live. We may not have solved the problem of poverty, but we certainly know a lot more about it than we did in 1964.

On other aspects of income inequality, some of the gaps in the data seem pretty preposterous. For example, federal taxes now constitute about 20 percent of gross national product and federal transfer payments to individuals in cash and in kind are running about 10 percent (U.S. Dept. of Commerce). Despite the clear potential of these instruments to alter income shares, there are no data regularly collected on the distribution of income after federal taxes; cash transfers to individuals are inadequately reported; and federal transfers in kind are not counted as income at all. It is quaintly anachronistic that the census counts as income the value of homegrown vegetables, but not the value to the recipient of medicare payments. Moreover, despite the interest of economists in the permanent income hypothesis and the obvious fact that it is hard to tell much about a family's well-being from one year's income, almost all income distribution data relate to single years and we are only beginning to make serious attempts to follow particular families for longer periods.

Why blame economists? Don't economists do the best they can to analyze the data that the government gives them? Yes, that's just the point. The passivity of economists about data collection is at least part of the reason why the right information rarely seems to be available for answering questions like: What's happening to the distribution of income? Disdain for data collection is built into the value and reward structure of our profession. Ingenious efforts to tease bits of evidence from unsuitable data are much applauded; designing instruments for collecting more appropriate information is generally considered hack work.

The Michigan Longitudinal Survey and the immense effort that has gone into income maintenance and other experiments connected with the concern for poverty may mark the beginnings of a new trend from passivity to activism in data collection by economists. But not much effort is yet going into maximizing the usefulness of obvious sources such as the census.

But enough of economists' shortcomings as data collectors. What do the available facts enable us to say about the distribution of income and how it is changing? The big news, of course, is that there is no news. Real incomes have been rising over the years at respectable if not spectacular rates, but measures of inequality have hardly changed at all since World War II. In a world in which change is normal—in which people are used to revolutions in technology and communications, racial and sexual upheaval, baby boom and baby bust, decline of cities and mushrooming of suburbs—the constancy of the size distribution of income stands out as remarkable.

The number of people living in poverty, of course, may be said to have declined as long as one associates poverty with a fixed level of real income. But if one defines the poverty line as, say, half the median income, one finds that the proportions in poverty have remained constant for a quarter of a century—19.0 percent of families had less than half the median income in 1947, 19.4 percent in 1972—while the absolute gap between the poor and the median family has widened. (See U.S. Bureau of the Census 1973b.)

The distribution of male earnings actually appears to be getting less equal, although most of the change relates to increases in the proportion of younger men in the labor force and in the frequency of part-time work by younger and older workers (Peter Henle). But it is the size distribution of family income that really seems to be stuck in a rut. The bottom 20 percent of families gets 5 percent of total income and the top 20 percent gets a little more than 40 percent—the numbers have hardly varied at all since the late 1940's. (See U.S. Bureau of the Census 1973a.) Whether one regards the distribution as acceptable or outrageous, one can hardly help being surprised that it remains so constant in the face of other changes, especially the growth of federal transfer programs whose benefits go mainly to low-income people. In-kind transfers do not count as income—if they did they would be slightly equalizing—but transfer income, such as social security and public assistance, does count. These payments have risen substantially in the last two decades and now amount to about $100 billion, one-third of the federal budget and about 9 percent of personal income (U.S. Treasury Dept.). One might well ask why they have not had a visible equalizing effect on the distribution of family income.

The answer appears to be that the preponderance of federal transfers has gone to retired persons and a small part to women heading families; that the recipients, especially old people, are better off, both absolutely and relatively than they used to be; but that the equalizing effect of all this on the income distribution is offset by a combination of two other factors. One is earlier retirement—only 24 percent of men aged 65 and over were in the labor force in 1972 compared with 46 percent in 1950 (U.S. Dept. of Labor). The other is the increased tendency of adults at all ages to head their own households.[1] Young people move out of the parental household sooner than they used to, old people are less likely to live with their children, and women—especially black women—are more likely to be family heads than they were a couple of decades ago. This increase in units, often at low-income levels, has made the distribution of income less equal than it would otherwise have been even though it doubtless reflects an increase in economic well-being. Rising incomes, including transfers, have enabled Americans to increase their consumption of a luxury good—the luxury of living apart from their relatives.

If one looks at the position of blacks in the income distribution the picture is a bit confusing. Since the early 1960's the ratio of black to white median earnings has risen somewhat for males and rather spectacularly for females.[2] The ratio of black to white family income, however, rose in the 1960's but has since declined (U.S. Bureau of the Census 1974). It is hard to see how these statements could both be true unless one remembers two other facts—that more black families are now headed by women and that a declining proportion of black men of all ages are in the labor force.

These various statistics reveal clearly that more is happening to the distribution of income than the constant size distribution would indicate—equalizing and unequalizing forces are cancelling each other out, although there is no obvious reason why they should continue to do so. The statistics also reveal something discouraging to economists—some of the most important influences on the size distribution of income are factors to which economists have given little attention. We do have some theory about the distribution of full-time male earnings, to which I will return presently, but we do not have much to say about family income. That depends on the groupings in which people choose to live (a subject that economists have always left entirely to God and the sociologists), the propensity of various family members to work (a subject that has evoked only a little interest from economists), the actions of the government, and the interactions of all these factors.

II. What Would Be Better?

Next we come to the question of norms or goals. What do economists have to say about the optimum distribution of income? Or at the very least, can we say anything about what would be better than the distribution that now obtains?

In a truly amazing act of self-denial, economists of the past generation have taken themselves out of the theoretical-philosophical discussion of how income ought to be distributed. Since they abandoned the notion that interpersonal comparisons of utility were possible and rejected the commonsense assertion that taking a dollar from a rich man and giving it to a poor man generally enhances the total welfare, they have not only stayed out of the argument, but devoted some of their most ingenious intellectual efforts to explaining to each other why they had so little to say.

---

[1] See U.S. Bureau of the Census, *U.S. Census of the Population, 1950,* "Marital Status"; *U.S. Census of the Population, 1960,* "Persons by Family Characteristics"; and *U.S. Census of the Population, 1970,* "Persons by Family Characteristics."

[2] See U.S. Bureau of the Census, *Current Population Reports,* "Consumer Income," various issues.

My own gut reaction is that the commonsense notion is correct—taking a dollar or a thousand dollars from a millionaire and giving it to a sharecropper with three hungry children *does* enhance total wellbeing. The interesting question is not whether this extreme proposition is true or not, but how far to carry it. The fact that members of our profession in the last decade have put hundreds of thousands of manhours into studying the very poor and designing policies to assist them attests to the fact that my gut reaction is widely shared, if not theoretically defensible. We are all pretty sure that raising the bottom of the income distribution enhances total welfare; we just aren't too sure what we think about redistribution in the middle.

The theoretically sanctioned position of the economists, however, biased their efforts away from redistribution and toward concern with economic growth. The bias is not necessarily detrimental to the cause of improving the share of the poor: it is probably easier to obtain political support for redistribution within a rapidly growing total than within a stagnant one, which may explain why there was so much more apparent support for helping the poor in the mid-1960's than there is now.

The one thing economists did feel that they could talk about without intellectual embarrassment was the tradeoff between equality and efficiency. One part of this argument involved the classical defense of the efficiency of paying all workers their marginal products, although everyone agreed it was not always possible to do this; another involved the labor supply effects of alternative tax rates, both positive and negative. The last question was the most relevant to actual policy debate since, except for minimum wages, the political decision makers seemed ill disposed to interfere with wage determination in an effort to equalize, but amenable to progressive taxes and income-tested transfers. Perhaps this was because economists were pushing in this direction. In any case, considerable effort has gone into estimating the possible labor supply costs of equalizing through taxes and transfers, but it seems only fair to say that the jury is still out.

There is one goal with which economists and a lot of other people have felt far more comfortable than they have felt with the goal of redistributing income in the present, namely, the goal of equalizing opportunity to earn income in the future. As long as equalizing opportunity was identified with lowering barriers to entry, such as racial or sexual barriers to education and training or to credit and job markets, or even when it was identified with equalizing productivity-enhancing public expenditures, such as expenditures for education or public health, economists and others felt they had something that was at least intuitively defensible. First, the current resource costs often appeared to be small. Letting blacks into unions or women into law school did not require any significant reallocation of current resources, although equalizing public expenditures often did. Second, the losers were not giving up income they had, only their expectation of future income gained by an unfairly advantageous starting position. Third, and most important, equalizing opportunity seemed likely to increase aggregate income, rather than reducing it. Conventional economic theory suggested that redistributing current income would reduce the size of the total—if one did it by paying people something other than their marginal product, there would be misallocations, and if one did it by progressive taxes and transfers, there might be less labor supplied—although persuasive empirical evidence to support the theory was lacking. It was not hard to show theoretically, however, that equalizing opportunity, by giving naturally productive people a chance rather than a handicap, might actually increase total output. Hence, equalizing opportunity to earn income, like free trade, was something the preponderance of the profession unabashedly viewed as a good thing. It was not quite a free lunch, but it came close.

Unfortunately, however, further thought and empirical research, much of it by noneconomists, has taken the almost free lunch off the table. It is now evident that lowering artificial barriers and equalizing public expenditures for human investment (for example, education expenditures) make relatively trivial contributions to equalizing opportunity to earn income. The family, it turns out, is an enormously powerful influence on individual capacity to perform successfully in school and in the job market (James S. Coleman et al.). One cannot discuss how far society ought to go in using public resources to compensate differences without facing the necessity from which economists are so eager to escape—that of balancing the welfare of the taxpayer and the beneficiary and, indeed, of trading the resources of this generation for more equal opportunities in the next.

At this point the economists are tempted to leave the field to the moral philosophers with the instruction: you guys tell us what ought to be done about equalizing incomes or opportunities and then we will help you do it. But unfortunately, there is no escape. Probably the most ambitious recent attempt by a moral philosopher to evolve a set of rules for economic justice is the widely discussed work of John Rawls. Rawls suggests that the way to decide what is fair is to think about what rules people would choose for a society in which they knew they were going to have to live, but had no idea what their social or economic status would be. He asserts, offering no empirical evidence, that people in this circumstance would be absolutely averse to risk. Taking no chances, they would agree on two fundamental rules. The first is that certain basic rights are to be distributed absolutely evenly and cannot be bartered away, either for individual or social benefits. The second has two parts—that income differences are tolerable only when they benefit everyone, including the least advantaged, and that advantageous positions are to be passed out on the basis of achievement with everyone having an equal chance at them.

These rules are appealing, but they do not get us any farther down the road to policy choices than the economists' agnosticism. In the first place someone has to decide what are the basic rights to which everyone has equal claim. If one includes what Arthur Okun has called "resource-using rights," such as the right to a minimum income or a decent house or access to medical care, then one has moved a long way toward income equalization, perhaps at such cost to the total product that the least advantaged are worse off (see Okun). If one defines basic rights narrowly, then the second proposition is troublesome. Deciding which inequalities are beneficial to the least advantaged is an awesome task for empirical research. At the moment, for example, the trickle-down school of fiscal policy is suggesting that the way to stimulate the economy is to lower taxes on business investment and capital gains, an apparent increase in inequality that is supposed to benefit the least advantaged by stimulating the demand for labor. Whether this school is in fact right that the least advantaged would benefit is an empirical, not a theoretical question. Moreover, as Coleman has pointed out, Rawls's endorsement of equality of opportunity is either trivial or involves very difficult decisions about what proportion of the national product should be devoted to overcoming the influence of the family on income-earning capacity and, indeed, about the extent to which the family should be perpetuated as a socializing force if it reduces equality of opportunity.

In my own view, discussions of optimum distributions of income or opportunity are fascinating intellectual parlor games unlikely to make any contribution to the policy debate on income shares. Economists who want to make themselves useful would be better advised to put their effort into defining some real policy alternatives, estimating their costs and effects and explaining what they are talking about to the public and its political representatives so that the political process, for all its faults, can at least operate with more knowledge of what is possible than it has at present.

It should be noted that the revealed preferences of the American public with respect to income-equalizing policies are a mixed bag, judging from the current array of federal activities that directly affect the distribution of income. The preponderance of federal cash transfers goes into programs whose effect is to even out income inequality over the life cycle (e.g., social security and other retirement programs) or to reduce transitory income inequality due to inadequate aggregate demand for labor. Few resources are devoted to offsetting differences in earning capacity as such. On the tax side, popular tolerance for the regressive payroll tax and lack of support for income tax reforms that would clearly benefit the majority of taxpayers suggest either that the political process is a poor reflection of majority preferences, or that no one understands what the choices are, or both. It is in the solution of this latter problem—elucidating the options—that economists seem to have the most to contribute. Before one can do that, however, one needs a workable theory of how income in fact gets distributed.

III. Why Things Are as They Are

It seems only reasonable that those concerned with the size distribution of income should ask economists why they think the distribution is so unequal.[3] Why do so many have so little while a few have so much?

Economists have made some valiant efforts to answer this question. Vilfredo Pareto even thought he had discovered a natural law that accounted for the shape of income distributions in a wide variety of places in various time periods. Pareto was concerned with the distribution of total income, but most of the other theorists who have tackled the problem have confined their attention to the distribution of earnings, especially full-time male earnings. Indeed, it is something of a paradox that what economists think of as "income distribution theory" relates to the distribution of earnings of fully employed males, while what they think of as "income distribution policy" means almost everything else.

Over the years various theorists have offered explanations of the puzzling fact that the abilities of individuals, whether physical or intellectual, are distributed normally while the earnings distribution is more nearly normal in the logs. This observation suggested that some multiplicative factor related ability and earnings; for example, that several different kinds of abilities determine earnings and that they are multiplicative in their effects, or that individual earnings, initially proportional to some index of ability, are subject to a series of random proportional shocks. These theories were intriguing, but devoid of any policy significance. Indeed, only within the last fifteen years has economics produced a theory—the human capital theory—that both offers a credible explanation of the shape of the earnings distribution and some potentially testable hypotheses about how one might intervene to change that distribution.

---

[3] For a good general review of income distribution theory, see Martin Bronfenbrenner and Jacob Mincer (Mar. 1970).

Human capital theory was appealing to scholars not only because it offered hope of accounting for the shape of the earnings distribution, but also because it accounted for part of the unexplained variation in earnings among individuals, regions, and subgroups in the population.[4] It even answered some questions that no one had thought needed asking, like why people tend to get their education when they are young.

But an even stronger appeal of the human capital approach was to people with an interest in policy and a predisposition to the belief that education was a good thing and ought to be better funded. (Economists tend to fall in this category, not only because they have invested so much in their own education, but because they so frequently earn their living educating the young.)

First, of course, the human capital approach provided a rationale for education spending in a growth-oriented society. The exact calculations were not important—no one cared much whether social rates of return on additional education were 5 percent or 15 percent as long as they seemed to be positive. If one was proeducation, one could imagine all manner of external benefits of education—from reduction of crime to more intelligent behavior in the voting booth—that had not been counted in the rates of return. The point was that education could be thought of as an investment and therefore a worthy use of resources, not a wasteful frivolity that consumers wanted only for their own enjoyment or for impressing their neighbors. This basic idea may have done as much as Sputnik to convince legislators of the 1950's and 1960's that increasing education spending (more than was dictated by population increases alone) was in the public interest.

Second, the human capital approach suggested that subsidies for education and training could be used to alter the income distribution in one of two distinct ways: (1) raising the relative position in the income distribution of people whose parents were near the low end; and (2) narrowing income differentials in the coming generation compared with what they would otherwise have been.

It is important to be clear that these are not the same thing. Improving the relative position of some small fraction of the next generation (i.e., their own children) is what motivates parents who scrimp and save to move into a neighborhood where the schools are better or to take a second job to send the kids to college. Parents are not trying to narrow income differentials among the next generation; on the contrary, they think large differences between the incomes of the well educated and the poorly educated are just fine as long as their own kids make it into one of the top groups. The motivation behind the education and training components of the 1960's war on poverty appears to have been similar to that of parents seeking to improve the chances of their kids. The notion was that additional public resources could be concentrated on a fairly small number of young people (or even adult workers) who would otherwise be likely to have low incomes, and that this additional investment would improve the relative position of the recipients. It might, of course, be more efficient to improve their relative position by giving them money, but the education route had more political sex appeal.

---

[4] See, for example, Gary S. Becker (both references), Mincer, and Barry R. Chiswick.

Success in this endeavor requires: (1) that the additional resources devoted to the education of the disadvantaged actually do some good, in the sense that the students or trainees actually increase their skills or acquire valuable credentials, and (2) that there not be so many people with the new skills or credentials that the wages they command fall and the income advantage diminishes.

Alternatively, however, the public policy objective might actually be to narrow disparities in the distribution of income among the next generation. Human capital theory suggests that one way to do this—again not necessarily the most efficient way—is to increase the proportion of people attaining higher levels of education and training. Other things being equal, this would increase the competition for jobs with high education or skill requirements and lower the wages those jobs command. At the same time, wages of jobs requiring less skill and training would tend to rise, and disparities in income among education and skill levels would be reduced.

Such allegations are hard to prove, but I suspect that narrowing disparities in income was *not* among the significant motivations for the major expansion in public subsidies for higher education that occurred in the last fifteen years nor for the education and training programs of the war on poverty. The proponents of these programs were focused on giving all young people including the children of the poor a chance to improve their expected income, *not* on reducing inequality per se. They hoped, in fact, that the income differentials between education groups would remain large rather than erode under the impact of the policies. There would then be no losers, only gainers.

One of the minor economic mysteries of the 1960's was that these hopes were not disappointed. Despite substantial increases in high school graduation rates and in the ratio of college to noncollege workers, income differences did not narrow at all. In 1960, for example, men aged 35–44 with four years of college had 1.91 times the median earnings of those with only eight years of schooling; in 1970 the comparable ratio was 1.92—another one of those "stuck" statistics.[5]

At the time an explanation seemed in order and several have been offered. First, technology may be changing in such a way that the demand for employees with the skills acquired in education is shifting upward at about the same rate as the supply. Second, labor markets may not actually work in the neoclassical way that human capital theorists have assumed they do. Alternate theories of labor market functioning abound at the moment and several have been used to explain why income differentials among income groups did not narrow in the 1960's. The most drastic explanation simply rejects the human capital theorists' contention that education enhances productivity, and suggests that the only reason that education is correlated with income is that the combination of ability, motivation, and personal habits that it takes to succeed in education happens to be the same combination that it takes to be a productive worker. If one takes this view, then changes in the supply of people with various amounts of education are simply irrelevant to the distribution of earnings and there is no reason to expect one to predict the other.

An intermediate explanation may be that those aspects of education that count in the labor market are not actually being equalized as much as the data on years of schooling would suggest. There may be a basic number of years of schooling that society expects everyone to have and that counts for nothing in the labor market because all but the subnormal have it. Much of the shift in education in recent years can be construed as a shifting upward of the base level, rather than a change in the proportions of workers having significant amounts of education beyond the base level.

---

[5] See U.S. Bureau of the Census, "Educational Attainment," 1963, and "Educational Attainment," 1973.

Whatever the merits of these hypotheses, more recent data suggest that the decline in the relative earnings of recent college graduates, expected in the 1960's, is actually occurring in the 1970's and that the delay may have been attributable to the fact that such a large fraction of the college graduates of the 1960's flowed into graduate and professional schools and did not hit the labor market immediately (Richard B. Freeman). It is too soon to tell, but if the drop in relative earnings of highly trained people does continue, it will be somewhat harder to use education as a tool for enhancing social mobility. It will be good news, however, for those who see education as a way of narrowing disparities in income.

Even if one accepts human capital theory and conventional views of labor market operation as, so far, the best explanation of the earnings distribution that we have, one cannot make a strong case for relying on educational investment as a primary tool for changing the income distribution. At best the effects of education are slow in appearing, the proportion of the variance in income explained by education is not large, and firm empirical evidence is weak that increased expenditure for education of particular students actually increases their future income. Alternate views of the labor market and alternate explanations of the correlation between income and education—such as the queuing, screening, and dual labor market models—offer even less support for education as a tool for increasing income equality.[6]

IV. What Can We Do About It?

The pay-dirt question, of course, is what policies might change the distribution of income—presumably in the direction of greater equality, since no one is likely to advocate less equality as an explicit objective. How would such policies work? What would they cost?

There has been a major accumulation of useful knowledge in this area in the last ten or fifteen years and economists have good reason to be proud of their contributions. They have put tremendous effort into the study of redistributional policies and have learned a great deal that ought both to improve politicians' abilities to make informed choices and the workability of any programs that are chosen. It is not economists' fault that the real world is so complicated and that much of what has been achieved is an elaborate documentation of the fact that there are no quick and easy solutions to problems that have been around for a long time.

It can be argued, of course, that it is better not to know so much, because knowledge of the complexity of the problem paralyzes action. But the history of the 1960's teaches the opposite lesson: naive hopes and hastily designed programs can produce a disillusionment with public intervention in general that is itself paralyzing.

So let us suppose that there is a new High Wizard in Washington—you may identify him as a President or a Congress or both—who has a landslide election behind him and an interest in fairness and in improving the lot and the future chances of those below the middle of the economic ladder. The Wizard calls his Trusty Economic Adviser (TA) and says: You and your colleagues have been working on this problem for a long time and I want to know what you have learned. What policy choices do we really have? Trusty Adviser, of course, starts by saying that he needs another ten years to work on the problem and a lot more money for research, but the Wizard won't buy that—he wants to know what is known *now.* OK, says TA, but first, tell me what you want to happen. Three things, replies the Wizard. First, I'd like the very poor to be better off—closer to the average. Second, it doesn't seem fair to me that there should be such concentration of income in the hands of the families at the top. And third, I think there ought to be more chance for people at the bottom to move up or at least to see their children move ahead. It's bad when people get stuck at the same level with no hope of improving their position. You want to do all three? asks TA. I'd at least like to make some progress, says the Wizard, but of course I have some constraints. The people who elected me are not socialists or anything. So just tell me what's possible without making the government very much bigger than it is now. What could be done with taxes and government programs? Let's start with education.

---

[6] See, for example, Kenneth J. Arrow and Lester C. Thurow and Robert E. B. Lucas.

Education may be a good thing to spend more resources on, says the TA, and there certainly won't be much upward mobility unless low-income children have access to education. But I have to warn you that increasing education spending is probably a slow and relatively ineffective way to change the income distribution. You could, of course, try spending substantially larger sums on the education of children from families in the lower part of the distribution. You could reach them earlier through preschool programs; you could buy them more teachers, better facilities, and special "compensatory" programs in elementary and secondary school. All this would make their lives a bit less grim, and in theory this would improve their performance in school and increase their chances of earning more money later on.

Why do you say "in theory"? interjects the Wizard. Because in practice no one knows how to do this. One thing that makes it difficult is that your national government does not control how local schools spend money, but the main problem is that no one knows what to tell the schools to do. Statistical studies do not prove that higher spending for schools is associated with better performance of children—as long as the children's home background is held constant—and evaluations of special new educational programs have not yet shown convincingly that any of them lead to permanent increases in the school performance of low-income children or any other children. Of course, a bigger research effort on teaching and learning might lead to a breakthrough that would make education spending more effective. Such a breakthrough might help low-income children more than others—the way control of contagious diseases helped the poor more than the rich—but there is also the risk that it might be of most help to those who need it least.

Are you saying, asks the puzzled Wizard, that spending more per child in elementary and secondary school is a waste of money? Not at all, replies TA. I'm just saying that economists cannot prove that the additional spending will raise children's test scores or incomes. But we can't prove that increments in most public expenditures have measurable effects especially on income. Good education is something most people want for themselves and their children. You politicians have to figure out how much they want it and how to distribute it fairly and I'm afraid you aren't going to get much help from economists.

But, says the persistent Wizard, how

about encouraging students from low-income families to stay in school longer, especially making it financially possible for them to go to college? We're on firmer ground here, admits TA. Making the college option available to students from all income levels is bound to increase mobility from one generation to the next. If a lot more students go on to higher levels of education, there may be more competition for high status jobs and the distribution of income may even become more equal. Of course, you may have to deal with some disappointed people. Some of those who thought a college education was going to guarantee them a professional or managerial job and a high income may feel they were misled. Also it is my duty as an economist to point out that the education purveyed by our colleges and universities is expensive and if your objectives are to raise the relative income of the poor or to narrow income differentials (not just to facilitate upward mobility in status or occupation) it might be cheaper to do it by taxing higher income people and supplementing the incomes of those who earn less.

Ah yes, says the Wizard, you economists developed something called a negative income tax to solve this problem. Should I try that?

You certainly ought to consider it, replies TA. A lot of economists got excited about the negative income tax in the early 1960's because it seemed to be such a clean, elegant solution to several problems at once. It would put a floor under everyone's income—indeed, the floor could bear some relation to average income, as you suggested—without destroying incentives to work. Combined with reform of the positive tax system, to raise average effective tax rates on high-income people, it could not only raise the share of those at the bottom of the income distribution, but could also reduce the concentration of income among those at the top. In the last decade we economists have done a lot of hard practical work on this general idea. Besides theoretical and statistical studies of the usual sort, we have designed and analyzed actual experiments and have helped lay out detailed plans for turning the government's welfare system into something resembling a negative income tax or having the tax authorities take on these new responsibilities.

That's nice, says the Wizard, a little impatiently, but what have you learned from all of this that will help me decide what to do? I was coming to that, says TA. First, the experiments proved that a negative income tax was not just a crackpot idea; it was actually possible to administer one. Moreover, male family heads did not quit working just because they were getting the payments, although some of their wives did. We still don't know, of course, how the results of a real long-term national program would differ from those of a short-term local experiment. Second, all the work on legislative proposals revealed that it was very difficult and expensive to use a negative income tax both to supplement the wages of low earners and to support families with no earners at all; that it was hard to start a national program when wage levels and public attitudes in various parts of the country were so different; and that a lot of complex tangles have to be straightened out before a negative income tax can be made to mesh with the positive tax system and other existing government activities.

Are you telling me, asks the Wizard, that economists don't favor a negative income tax anymore? No, no, replies TA patiently, lots of them, including economists of very different political persuasions, still think that a negative income tax, plus reform and simplification of the positive tax system, is the best approach to reducing the concentration of income and improving the position of the poor. It's just that we now realize that a negative income tax is no easy panacea, and we are increasingly skeptical of our ability to explain it to politicians and the public and persuade them it is a good idea. Some of us tried that in a recent presidential campaign and it didn't work out too well. Hence, many of us are leaning to more eclectic solutions.

Tell me what you mean by that, demands the Wizard. Well, a lot of economists used to be purists about redistributing income in cash so that recipients of the benefits could choose freely how to spend them—presumably buying what they needed most. But some of us have realized that the public may be more willing to pay taxes to ensure that everyone has the basic essentials, such as food, housing, and medical care, than to hand over cash to people whose spending priorities may differ from their own (Thurow and James Tobin). So you may be more successful with your redistribution program if it features health insurance and housing assistance, rather than just a negative income tax. Of course, there are problems with this too. If people are eligible for a lot of different benefits, each of which is related to their income, they may in effect be facing very high marginal tax rates on their income. There is also a problem of deciding what form assistance should take. Most economists believe in markets so they usually favor giving the equivalent of a voucher to the consumer and letting him choose among suppliers; but, of course, suppliers like it better when the subsidy goes directly to them.

Actually, if you are serious about your objectives, you might try all these policies at once: education, negative income tax and positive tax reform, plus greater efforts to ensure everyone such essentials as medical care. Is that a recommendation? asks the Wizard. Oh, I don't make recommendations, replies TA, I just tell you what the choices are.

This facetious-sounding conversation comes close to mirroring, as I perceive it, what the TA's have to tell the Wizards right now. Of course, the contributors of advice include only those who hold some hope for change within the system and do not view all government actions as part of a conspiracy of the few to oppress the many. Moreover, at the moment most of the Wizards are not listening.

Clearly, economists have done a lot to make the policy discussion more informed and constructive than it was ten or even five years ago. We may have contributed to the loss of innocence, but we have also accumulated some practical knowledge that can be used to effect change if a political Wizard with a passion for fairness and a will to change the income distribution asks for advice.

Actually, however, Wizards don't ask—they generally have to be persuaded that a problem exists whose solution would be popular, at least with substantial numbers of the electorate, and, more important, that change is possible. Once that happens they may even pick up a half-formed policy proposal and run with it too hard and too fast. Although some economists may bemoan the lack of influence of our breed, my own view is that in the income distribution area the political world grabbed some economic hypotheses—such as human investment as a cure for low earnings or negative income tax as a solution to income inadequacy—before the economists were quite ready to turn them into proposed laws. The result was disillusionment that turned people away from the objectives, rather than causing them to renew their efforts to find solutions. If the new eclecticism among economists means, among other things, a new willingness to work hard on the details and possible limitations of policy proposals, it may presage an end to the enthusiasm-disillusionment cycle of recent years and some real, albeit incremental, movement in the direction of a more equal income distribution.

REFERENCES

K. J. Arrow, "Higher Education as a Filter," *J. of Pub. Econ.*, 1973, *2*.

G. S. Becker, *Human Capital and the Personal Distribution of Income: An Analytical Approach*, Woytinsky Lecture No. 1, Ann Arbor 1967.

———, *Human Capital*, Nat. Bur. of Econ. Res., 4th printing, New York 1972.

M. Bronfenbrenner, *Income Distribution Theory*, 2d printing, Chicago 1970.

J. S. Coleman, "Inequality, Sociology, and Moral Philosophy," Univ. of Chicago, unpublished.

——— et al., *Equality of Educational Opportunity*, U.S. Dept. of Health, Education, and Welfare, Office of Education, Washington 1966.

B. R. Chiswick, *Income Inequality: Regional Analysis within a Human Capital Framework*, Nat. Bur. of Econ. Res., New York 1974.

R. B. Freeman, "Overinvesting in College Training," dis. pap. 371, Harvard Instit. of Econ. Res., Cambridge, Mass., July 1974.

P. Henle, "Exploring the Distribution of Earned Income," *Monthly Labor Rev.*, Dec. 1972, *95*.

J. Mincer, "The Distribution of Labor Incomes: A Survey with Special Reference to the Human Capital Approach," *J. of Econ. Lit.*, Mar. 1970, *8*, 1–26.

———, *Schooling, Experience, and Earnings*, Nat. Bur. of Econ. Res., New York 1974.

A. Okun, *Equality and Efficiency: The Big Tradeoff*, Washington, forthcoming 1975, ch. 1.

V. Pareto, *Manual of Political Economy*, Clifton 1971, *7*, 11–31.

J. Rawls, *A Theory of Justice*, Cambridge, Mass. 1971.

L. C. Thurow, "Cash Versus In-Kind Transfers," *Amer. Econ. Rev. Proc.*, 1974, *64*, 190–96.

——— and R. E. B. Lucas, *The American Distribution of Income: A Structural Problem*, U.S. Congress, Joint Econ. Comm., Mar. 17, 1972, Washington 1972.

J. Tobin, "On Limiting the Domain of Inequality," *J. of Law and Econ.*, Oct. 1970, *13*.

U.S. Bureau of the Census, *U.S. Census of the Population, 1950*, Special Reports, Part 2, Ch. D, "Marital Status (1953)," Tables 1 and 2, Washington, *4*.

———, *U.S. Census of the Population, 1960*, Subject Reports, "Persons by Family Characteristics," Final Report, PC(2) 4B 1964, Table 2, Washington.

———, *U.S. Census of the Population, 1960*, Subject Reports, "Educational Attainment," Final Report PC(2) 5B 1963, Table 6, Washington.

———, *U.S. Census of the Population, 1970*, Subject Reports, "Educational Attainment," Final Report PC(2) 5B 1973, Table 7, Washington.

———, *U.S. Census of the Population, 1970*, Subject Reports, "Persons by Family Characteristics," Final Report PC(2) 4B 1973, Table 2, Washington.

———, *Current Population Reports*, Series P–60, "Consumer Income," various issues, Washington.

———, *Current Population Reports*, Series P–60, no. 90, "Money Income in 1972 of Families and Persons in the United States," Table 16, Washington 1973a.

———, *Current Population Reports*, Series P–60, no. 87, "Money Income in 1972 of Families and Persons in the United States," advanced data, Table 1, Washington 1973b.

———, *Current Population Reports*, Special Studies, Series P–23, no. 48, "The Social and Economic Status of the Black Population in the United States, 1973," Table 6, Washington 1974.

U.S. Dept. of Commerce, *Survey of Current Business*, Tables 1 and 13, Dec. 1974, *53*, no. 12, Washington 1974.

U.S. Dept. of Labor, *Handbook of Labor Statistics 1973*, Tables 4 and 7, Washington 1973.

U.S. Treasury Dept., *The Budget of the United States Government, Fiscal Year 1974*, Washington 1973.

THE ECONOMICS OF LABOR: AN ASSESSMENT OF RECENT RESEARCH

The Challenge of Dual and Radical Theories of the Labor Market to Orthodox Theory

By GLEN G. CAIN*

Ideally, a paper that attempts to evaluate a challenge to established theory should accomplish the following objectives:

1. Present the theoretical substance of the challenging theory, which is here called the dual and radical theories, and indicate how it differs from the existing theory, which I will refer to as neoclassical or orthodox theory;
2. Point out the empirical implications of the new theories and explain their differences with the existing theory;
3. Assess the empirical basis for discriminating between the competing hypotheses or, if this is lacking, provide the theoretical or evidential counterarguments of the neoclassical response to the dual and radical challenge; and, finally,
4. Spell out the policy implications of the competing theories.

Meeting these objectives is an impossible task for a short paper. The space constraint is compounded by the fact that in my judgment the dual and radical theories are too varied, incomplete, and amorphous to present concisely. The strategy I adopt to describe these theories in the longer paper consists of, first, discussing their linkages to historical criticisms of classical and neoclassical theory; second, developing their empirical challenges to orthodox theory. Only the second part is emphasized below.

To establish the context of the challenge, let us begin with the observation of Leo Rogin (p. 13) that: "new systems (of economic doctrines) first emerge in the guise of arguments in the context of social reform." The dual and radical (or *D–R*, for short) theories began to emerge in the 1960's when the movement for social reform mainly involved the war on poverty and the drive for full participation in the economy by minority groups, including women. Dissatisfaction with the pace and progress of reform in these areas and dissatisfaction with the conventional analysis of the problems and their remedies have led to arguments within the economics profession, especially on the part of the younger labor economists.

Let me simply assert, without defending the proposition as I do in the longer paper, that the neoclassical school does dominate the profession today, even within labor economics where, as I mention below, it was not dominant during the 1940's and 1950's. I will also assert that neoclassical theory means, in general: (1) the theory of the demand for labor contained in the marginal productivity theory and the associated research on production functions; and (2) the theory of labor supply consisting mainly of models involving the labor/leisure choice and human capital investments.

---

* Department of economics and Institute for Research on Poverty, University of Wisconsin. This is an abbreviated version of a longer paper, which is available from the Institute for Research on Poverty, University of Wisconsin, Madison, Wisconsin 53706. The longer version contains a bibliography, which is omitted here. I am indebted to a number of people at the U.S. Department of Labor, especially Fred Siskind, and at the Institute for Research on Poverty for their support. They are not responsible for errors, nor do they necessarily agree with the interpretations expressed in the report. I am grateful to Marc P. Freiman for his extensive assistance.

Theory, however, operates on at least two levels. As a framework for analysis, neoclassical theory implies a set of methods and techniques—for example, marginal analysis of behavioral relations in which income and prices are key variables. Second, as a body of substantive behavioral propositions, neoclassical theory implies a large number of qualitative predictions and quantitative estimates of parameters about behavioral relationships. The dual and radical critics attack both the methodology and the collection of predictions and substantive hypotheses. Those attacks inevitably spill over into policy disputes.

I. A Classification of Issues Raised in the Dual and Radical Literature

Out of this background we can discern three types of issues raised in the *D–R* literature: first, criticisms of the outcomes and processes taking place in the labor market; second, a rejection of the existing theories and methods that have purported to explain the operations of the labor market; and third, a call for new and more radical policies.

A. Empirical Generalizations

It is useful to distinguish between two types of facts which contribute to controversy. The first are those that indicate some sort of hardship or distress, such as high levels of unemployment and poverty. The facts about these problems may not be in dispute. The second type of descriptive empirical findings consists of unresolved and inadequately treated problems, which may or may not concern matters of hardship. The empirical challenges raised by the *D–R* spokesmen usually concern the overlap—empirical anomalies about pressing social problems. I will list eight problem areas, some of which have subtopics.

1. As noted above, the most important problem motivating the *D–R* economists was the persistence of poverty in conjunction with the affluence of society and the scope of governmental efforts to eradicate poverty. This problem is too broad, however, and we must turn to its more particular aspects, listed below as items 2 through 7.

2. One such problem is the failure of education and training programs which, according to the *D–R* critics, demonstrate a failure of the human capital models and call into question the conventional acceptance of a causal connection between formal education and labor productivity. I comment on the alleged failure of social action programs below. Concerning the general education-productivity relation, my judgment is that its empirical support is shakiest regarding the intensive margin of educational investment—that is, the effect of more educational resources, holding years of schooling constant. There is, I believe, ample empirical support for a payoff at the extensive margin—that is, positive returns to more years of schooling. The orthodox interpretation of this empirical relation is, however, called into question by the next point.

3. Employers used educational and training indicators merely as screening devices—either to discriminate on grounds other than productivity, or to serve as a proxy for potential productivity that is not directly related to what was actually

learned in the educational institution. The latter purpose, it should be noted, redefines the investment nature of education as "information"; it does not deny that any investment takes place. However, a denial of either the real productivity of education or of the potency of the informational content of educational attainments does challenge the orthodox assumptions about rational behavior and/or about the degree of competitiveness in the economy.

4. The income distribution remains as unequal now as it did twenty years ago, in spite of the decrease in the variance of educational attainments in the adult population.

5. A collection of problems concern the large and persistent differentials in earnings between white and black males and between males and females. These challenge neoclassical models of discrimination which predict a withering away of the differential under conditions of competition. Several more specific empirical puzzles in connection with labor market discrimination may be listed: the decline in black male labor-force participation rates relative to white males in recent years; the near-constant ratio of black-to-white earnings from 1950 to 1966 or so; the long-standing empirical relation showing a decline in the ratio of black-to-white male income with higher educational attainment; the flat wage-earnings profile of black males and all females relative to white males as revealed by cross-section surveys; and, finally, the stagnant trend in earnings and occupational attainments of women relative to men during recent years. Associated with this trend is the slower growth in female educational attainment compared with men from 1940 to 1973, despite the sharp increase in labor-force participation of women during this period.

6. The roles of monopolies, unions, and other sources of protected labor markets have been a long-standing problem area. The *D–R* spokesmen imply that these are so pervasive as to make untenable the empirical work of orthodox labor economists, which usually assumes competition.

7. Another area is the role of psychological variables measuring the attitudes and motivations of workers. Conventional economists have customarily viewed tastes for work as exogenous and as one of the (unexplored) causal variables explaining employment, earnings, and occupational achievements. The challenge of the *D–R* theories lies in their claim that tastes are endogenous and a result of one's labor market achievements.

8. Alienation among American workers is the final listed problem. If we assume as true the disputed contention that alienation and dissatisfaction among *U.S.* workers are widespread and increasing, the challenge to orthodox theory may be stated as follows: Why has the market not responded to the workers' tastes and preferences either by a redesign of jobs and upgrading of working conditions or simply by appropriate wage rate compensations?

The foregoing list of empirical generalizations about outcomes and operations of the labor market covers, I believe, the main bill of particulars in the *D–R* indictment. Let us now comment briefly on their theories of how the labor market operates.

B. Alternative Theories

In the absence of any single, well-articulated theory in the dual and radical literature, I will mention three versions that veer progressively further from orthodox theory.

Perhaps closest to the orthodox is Lester C. Thurow's job competition theory, which is contrasted with the orthodox wage competition theory (Thurow and Thurow and R. E. B. Lucas). Its main elements are: (1) the number and type of job slots are technologically determined; (2) the workers' skills (i.e., human capital) and wage offers are nearly irrelevant in determining the number and type of job positions actually filled; (3) queues of workers for jobs offered at fixed wages constitute the supply of labor, and the employer's assessment of the workers' trainability and adaptability determines which workers are hired; and (4) fluctuations in macro policies will lead to changes in the demand for labor and thus to changes in the lengths of the queues. The theory emphasizes the within-firm (or internal) labor market as the locus of decisions about allocations, promotions, and on-the-job training—all of which are relatively insulated from the external labor market. In many respects this theory is similar to the dual market theory.

The dual labor market theory of Peter B. Doeringer and Michael J. Piore depicts a labor market divided into primary and secondary markets. The former contains the jobs in large or profitable firms and the unionized jobs. These jobs are higher paying, offer promotion possibilities, better working conditions, and greater stability. The secondary labor market contains the low-paying jobs which are held by workers who are discriminated against and who have unstable working patterns. The discussion of this duality tends to be merely descriptive, and perhaps descriptive only of polar cases. Many of the theoretical ideas are similar to the job competition theory; namely, the demand-determined allocation of jobs and the downgrading of human capital characteristics as determinants of wage levels. In addition, considerable attention is given to the role of employer discrimination. Also, the worker's attitudes, motivations, and work habits are viewed as mutually reinforcing determinants of a worker's assignment (and confinement) to the primary or secondary sector of the labor market.

The third version of a labor market theory is referred to variously as radical, segmented, and stratified theory (see David M. Gordon). It expresses a more explicit critique of capitalism, acknowledges its ties to Marxian dialectical analysis, and emphasizes class conflicts. The dual labor market idea is sometimes expressed in terms of an analogy with an underdeveloped economy or even with a colony which is exploited by the primary economy. Radical theories are also similar to dual market theories in drawing upon sociological analysis of institutional change and power relations and upon psychological analyses of the determination of workers' and employers' attitudes, preferences, and motivations.

The expositions of these theories are stronger in their criticisms of neoclassical theory than they are in advancing a coherent replacement. Criticism of classical and neoclassical theory has a long and, in many instances, a distinguished history, so casting the *D–R* writings in this mold can be complimentary. To the extent that the issues raised have appeared before (even though in different terms) and remain unresolved, the challenge is all the more compelling. On the other hand, if the issues have been satisfactorily answered before, the challenge is less compelling. In either case an historical perspective can be informative.

Space limitations prohibit this development, but mention should be made of the similarities of the dual and segmented market theories to the Mill-Cairnes theory of noncompeting groups and to several ideas of the neoinstitutionalist group of labor economists who dominated their field in the 1940's and 1950's. Of this latter group, John Dunlop had written of internal labor markets, and Clark Kerr had advanced the idea of labor market segmentation in his famous article about the "balkanization" of labor markets.

C. Policy Implications of the D–R Theories

One set of policies advocated in the *D–R* literature concentrates on the labor market; another deals with the larger issues of power relationships and nonlabor market institutions in society. The first set of policies is most clearly distinguished by a focus on the demand side of the labor market. Specifically, public employment, wage subsidy, and antidiscrimination programs are stressed. Intervention on the supply side of the market, particularly the human capital investment programs of education, training, and job search assistance, is deemphasized if not rejected.

A second set of policies advocated by the radical economists is less specific. These include prelabor market conditioning of the consciousness of people, perhaps calling for a reorganization of schools and other community institutions. They may also advocate that workers gain a more dominant role in governing their work—thereby combating alienation and partly achieving a general realignment of political power.

II. The Modern Neoclassical Response to the D–R Challenges

Again, space limitations prohibit much more than a listing of comments. I have divided these into a methodological theoretical section and an empirical theoretical section.

A. Methodological and Theoretical Issues

The defense of neoclassical research includes the following points: (1) adherence to the view that positive economic analysis is separable from normative analysis and, in particular, a denial that neoclassical models presume a regime of harmony rather than conflict among various economic groups in society; (2) belief in the utility of building models which assume either that tastes and institutions are fixed and/or that these variables may be omitted from the model without biasing the estimated economic parameters, such as price and income effects; and (3) finally, a denial that neoclassical theories of wages and employment rest solely, as *D–R* spokesmen often claim, on the marginal productivity theory.

B. Empirical and Theoretical Issues

What follows is a listing of five topics which provide an arena for debate. My main message is that more empirical evidence is needed, but that the *D–R* position is generally weaker than the neoclassical position on a priori grounds.

1. Two questions arise about the ***occupational structure.*** Does a static snapshot reveal a duality, and if so, what metric of occupational status produces this bimodality?[1] Another question concerns occupational mobility. Over time, are workers with a definable premarket characteristic confined to one segment of the occupational spectrum? There is little evidence on this question.

2. The neoclassical theories of ***discrimination*** have developed along three lines: competitive models which predict a long-run disappearance of wage (or job) discrimination; noncompetitive theories, which are downplayed on grounds that monopoly and monopsony are not empirically important; and competitive-statistical models, which in my opinion place too much of a burden on informational unreliability to rationalize the extent of discrimination that we observe. In their empirical work, neoclassical economists have stressed prelabor market sources of the disadvantages for both blacks and

---

[1] Occupational segmentation associated with sex and race groups is treated separately below under the topic of discrimination. Presumably, the segmentation of teenagers and the aged could be set aside as well, since there is no puzzle about their occupational differences compared with prime age workers.

women, thus attempting to reduce the burden of explaining market discrimination. These approaches, taken in total and despite their defects, appear more persuasive to me than the *D–R* models of exploitation, employer strategies of dividing and conquering their work force, or self-perpetuating and vicious circle theories.

3. Two points may be made in response to the *D–R* emphasis on *unemployment and job instability*. One is that if the human capital models explain individual workers' low wages, then they are in principle capable of explaining other undesirable forms of job rewards, such as instability. Second, a more direct analysis of job instability is addressed in the job search literature of neoclassical economists, which also attempts to reconcile macro and micro theories of unemployment. This effort is just beginning, however, and there are some theoretical and empirical gaps. (See James Tobin.)

4. The question that separates the *D–R* from the neoclassical views of the issue of *wage rigidities and protected labor markets* is almost entirely one of the empirical frequency, duration, and importance of these market imperfections. The neoclassical economists have, in fact, used these ideas in studying underdeveloped economies and in analyzing minimum wage laws and the effects of unions on wages and employment in the *U.S.* economy.

5. The alleged *failure of training and educational programs* raises the messy problem of evaluating the evaluations of social action programs. I question whether a simple verdict of failure or success is justified. There is, however, one technical point about the statistical estimation of human capital models which has led to a misinterpretation by the *D–R* spokesmen. The allegation that human capital variables have small or zero effects on earnings of the lower strata of workers is often based on regression models in which the dependent variable is effectively truncated—consisting of predominantly low values. However, the regression coefficients of independent variables like years of schooling or training are clearly biased toward zero in these truncated regressions, so the pessimistic findings are partly artifactual.

III. Conclusion

My brief summary judgment of the *D–R* challenge is that it does not begin to offer a theory of the labor market that can replace neoclassical theory, despite our various degrees of dissatisfaction with the empirical corpus of that theory. The *D–R*'s theoretical and methodological criticisms of the neoclassical theory are not substantial and are often misguided; nevertheless, a tradition of criticism of orthodox economics is sustained and this is healthy. Neoclassical research can become terribly inbred and out of touch with policymakers or practical users of economic predictions. This danger is particularly acute because the standards for empirical verification are so weak.

The *D–R*'s theoretical contributions, which amount to modifications and additions to orthodox theory, are: (1) the ideas of the endogenous determination of attitudinal variables among workers; and (2) the institutional dimensions of internal labor markets—which enrich our understanding of the economics of bureaucratic organizations.

In the areas of empirical research and policy prescriptions, the dual and radical school represents an important voice. Although their research suffers, in my opinion, because it is not anchored to as tight and consistent a theory as neoclassical theory, this fault is compensated by their new ideas and their discovery of empirical anomalies in the orthodox paradigm.

REFERENCES

P. B. Doeringer and M. J. Piore, *Internal*

Labor Markets and Manpower Analysis, Lexington 1971.

J. Dunlop, "The Task of Contemporary Wage Theory," in G. W. Taylor and F. C. Pierson, eds., *New Concepts in Wage Discrimination,* New York 1957.

D. M. Gordon, *Theories of Poverty and Underemployment,* Lexington, 1972.

C. Kerr, "The Balkanization of Labor Markets," in E. W. Bakke et al., *Labor Mobility and Economic Opportunity,* New York 1954, 92–110.

L. Rogin, *The Meaning and Validity of Economic Theory,* New York 1956.

L. C. Thurow, "Education and Income," *The Public Interest,* Summer 1972, 66–81.

——— and R. E. B. Lucas, "The American Distribution of Income: A Structural Problem," committee print, U.S. Congress, Joint Econ. Comm., Mar. 17, 1972.

J. Tobin, "Inflation and Unemployment," *Amer. Econ. Rev.*, Mar. 1972, *62*, 1–18.

Economic Analysis of Trade Unionism

By George E. Johnson*

The study of the behavior and effects of trade unions is not currently one of the major growth industries of the economics profession. For example, the average proportion of all articles in the *American Economic Review* (*AER*), *Journal of Political Economy* (*JPE*), and *Quarterly Journal of Economics* (*QJE*) dealing with questions concerning unionism, which seems a reasonable index of the degree to which economists at large consider the study of unionism to be interesting and important, has declined steadily for the past thirty years. In the 1940's, 9.2 percent of the articles in these journals dealt with union topics. This fell to 5.1 percent in the 1950's, further to 2.3 percent in the 1960's, and to just 0.4 percent during the first four years of the 1970's. Clearly, this index is not going to fall very much more.

It is not difficult to understand why the relative interest in trade unions has declined. First, the fraction of the civilian work force organized by unions has remained roughly constant at about 25 percent since the early 1950's, and with the exception of government employees, there appears to be little likelihood of a major increase in the extent of organization in the foreseeable future. Second, unionism in its present form has been quite secure since the early 1950's; its existence is not a serious matter of public policy as it was prior to that time. Hence, there is little call (i.e., no funding) for a judgment on the question of whether unionism is a good or bad thing. Third, there has been a great deal of activity with respect to policies concerning discrimination, antipoverty programs, income maintenance, and the like, and this has further diverted attention away from unionism. Indeed, the proportion of all articles in the *AER*, *JPE*, and *QJE* which deal with labor economics topics other than unionism increased from 4.8 percent in the 1950's to 7.4 percent in the 1960's and has reached 10.3 percent in the first part of the 1970's. Fourth, the earlier labor economics placed a strong emphasis on the examination of institutions in themselves rather than on their effect on economic variables, and its methodology was rather different from that of general economics. The resultant theory of wages and income distribution was not very satisfactory and, perhaps more importantly, could not be readily integrated into the mainstream of the discipline. Because of the development of human capital theory, we now have a much more satisfactory theory of income distribution, but the analysis of the effects of various institutional rigidities, unionism being one of the major ones, is not well developed. As D. Hamermesh (1973) suggested, it might be optimal for some of the resources presently devoted to the refinement of the new labor economics to be reallocated to a consideration of some of the problems considered by the old labor economics. As my survey of a subset of the latter will suggest, there remain a good number of questions which have not been resolved.

In the remainder of this paper I shall discuss—somewhat selectively—recent progress which has been made in answering three questions concerning the effects of trade unions. The space constraint has ruled out a similar discussion concerning behavioral questions, but it is worth noting that the problem of modeling trade union behavior has proved to be virtually intractable. This is because (1) there is no consensus on the goals of union activity (see W. N. Atherton) such as exists with respect to the firm or consumer and (2) the received pure theory of bargaining is devoid of operational content. The absence of a solid theoretical foundation has handicapped the economic analysis of trade unions and has surely contributed to its decline in relative attention.

---

* The University of Michigan.

I. Relative Wage Effects of Unionism

Regardless of what unions seek to accomplish or how they bargain with management, the question of how much more union members receive than comparable nonunion labor is both interesting and resolvable, and it has received a great deal of attention in the past few years. The stimulus to this recent research was the work of H. G. Lewis (1936b) in which the methodology of the topic was developed, all major previous studies were dissected and extended, and new sets of estimates were derived. Lewis' estimate of the proportionate wage advantage of union members over labor in the nonunion sector was 15 percent in the 1955–1958 period, but the estimate ranged from essentially zero in the late 1940's to 38 percent in the early 1930's,

The method of estimating the effect of unions on the relative wages of their members is upon first consideration fairly straightforward. In principle the wage rate for the ith worker (or the average wage for the ith industry or occupation) could be estimated by an equation of the form

$$(1) \qquad \log W_i = \beta U_i + \gamma X_i + \epsilon_i,$$

where W_i is the wage, U_i is a one-zero variable denoting union membership status (or the average extent of union membership for aggregate data), and X_i is a vector of all other variables besides unionism which influence wages (education, sex, race, experience and its square, geographic location, and the like), and ϵ_i is a random disturbance term. The estimated value of β provides an indication of the effect of unionism on relative wages. In particular, the proportionate wage advantage due to union membership, other things equal, is $\lambda = \exp(\beta) - 1$. The problem is to obtain the appropriate set of adjustment variables so that the estimator of β is unbiased. The early studies as well as Lewis' were based on data which did not permit much adjustment for other factors, but the development of micro data sets and computing machinery gave rise to a number of studies which could control more carefully for variations in (at least most of) the elements of X_i. L. Weiss used the 1960 Census 1/1000 tapes with industrywide unionization data and obtained results implying an average estimate of λ of about 20 percent for all male operatives and craftsmen in unregulated industries. F. Stafford employed Survey of Consumer Finances data for 1966, which had information on individual union membership status, and obtained estimates of λ of between 18 and 52 percent for all wage and salary workers. A. Throop used aggregate industry data and regressed wages on unionism, a skill index, and average city size of residence of the employees of each industry. His estimates of λ were 25 percent in 1950 and 30 percent in 1960.

The wide range of estimates of the union relative wage effect was consistent with Lewis' finding that λ varies over time and across industries. Thus, a number of studies attempted to disaggregate the data to obtain estimates of λ for various subsets of workers. In terms of equation (1), this means that U_i influences the dependent variable interactively with some of the elements of X_i. The Weiss study attempted to estimate the differential effect of unionism for different degrees of industry concentration. S. Rosen developed a theoretical model in which the threat effect of union organization causes employers of nonunion labor to take account of the union wage level in setting their wage offers. He then estimated values of λ of 38 percent in highly organized industries and of only 10 percent in lightly organized industries. The Stafford study found the highest union effect for the lowest occupational categories and the lowest effect for white-collar workers. G. Johnson and K. Youmans found that λ declines with education and is greater for the youngest and oldest age categories than for the middle-age categories. O. Ashenfelter estimated that unions increase the wages of black males by 21 percent as compared to 10 percent for white males. M. J. Boskin attempted to provide estimates of λ for all separate occupation-region-race-sex groups, but as R. Oaxaca has shown, its methodology was in error.

These studies, as well as others I have not mentioned, are all attempts to refine the basic model in order to arrive at *the* estimate (or set of estimates) of the relative wage impact of unionism. I am, however, a trifle skeptical about the appropriateness of this model because of its assumption that U_i is exogenous—although my skepticism is *not* widely shared by other labor economists. Consider the question of why wage differences net of quality and other factors might exist in the absence of unionism. Firms with production processes which are characterized by relatively high specific training requirements will attempt to maintain a low rate of labor turnover by paying higher wages than average; firms with lower training requirements, on the other hand, will maximize profits by paying lower than average wages (see J. Pencavel). Now workers in the high-wage firms will have a greater interest in the conditions of their employment than workers in low-wage firms, for the former are aware that they are unlikely to improve their lot by engaging in job search. One of the major functions of unionism is to regulate the nonwage conditions of employment for their members—including speed of production, grievance procedures, and seniority systems (see S. Slichter, ch. 2, and A. Alchian and H. Demsetz, p. 790). Thus, workers in what normally would be high-wage firms are more likely to demand unionism, and union organizers are more likely to supply their services to workers who are characterized by stable employment. This scenario suggests, holding labor quality and other factors constant, that for interindustry data such as Throop's (a) the wage level depends on unionism and the degree of specific training requirements, (b) the propensity to quit depends negatively on the wage level, and (c) unionism depends negatively on the propensity to quit. It is a straightforward matter to show that the ordinary least squares estimator of the effect of unionism on wages which does not both include the specific training requirements variable and take account of the joint determination of wages, the quit rate, and unionism will be biased in an upward direction. For micro data sets, the above argument suggests that union workers happen to be located in jobs which would have paid higher than predicted wages (on the basis of the included elements of X_i) even in the absence of unionism. In this case, the estimator of β will receive some of the credit which should go to the omitted specific training variable.

There are other approaches to the problem of the endogenicity of union membership (M. Reder, Ashenfelter and Johnson, and H. Levinson), but I will not go into detail at this time. The chief implication of this is that there is some cause to be skeptical concerning the accuracy of previous estimates of λ. What is needed to resolve this issue (and others not concerned with unionism) are data on individuals gathered jointly from households and establishments.

II. The Impact of Unionism on the Inequality of Income

A positive relative wage effect of unionism will have two effects on the personal distribution of income and welfare. First, the larger the union/nonunion relative wage the smaller will be employment in unionized relative to nonunion industries. Lewis (1963a) obtained an estimate of the elasticity of relative union/nonunion employment with respect to the relative wage of about minus one. This means that an increase in λ of five percentage points will, given an initial extent of organization of 25 percent, decrease union sector employment by (roughly) 3.8 percent and increase nonunion sector employment by 1.3 percent. H. G. Johnson and P. Mieszkowski used a two-sector general equilibrium model to investigate the effects of unionism on income distribution. After plugging in plausible values of the various parameters, they conclude that "most, if not all, of the gains of union labor are made at the expense of nonunionized workers, and not at the expense of earnings on capital." In addition, they calculate that the loss in aggregate welfare due to the inefficiency in the allocation of resources caused by unionism is only .33 percent (at λ equal to 15 percent).

Second, holding the size of the union and nonunion sectors constant, variations in λ will obviously influence the distribution of earnings. Whether the degree of income inequality is increased or decreased due to unionism depends on where union members would have placed in the distribution of earnings in the absence of unionism. Treating X_i in equation (1) as a composite variable (positively) influencing wages, the variance of the logarithm of individual wage rates is

$$(2)\quad \mathrm{var}\,(\log W_i) = \beta^2\,\mathrm{var}\,(U_i) + \gamma^2\,\mathrm{var}\,(X_i) + 2\beta\gamma\,\mathrm{cov}\,(U_i, X_i).$$

Unions increase (decrease) income inequality as $2\gamma b_{X\cdot U} \gtrless -\beta$, where $b_{X\cdot U}$ is the slope coefficient in a regression of X_i on U_i. In words, this means that in order for unionism to decrease income inequality it must be true that union members would, in the absence of unionism, be fairly well toward the lower end of the income distribution. This is, in fact, not the case. Union members generally possess characteristics which would put them in the middle of the income distribution, which means that $\mathrm{cov}(U_i, X_i)=0$, so that unionism probably has a *slight* disequalizing effect on the distribution of income (see A. Rees, pp. 98–99, and A. Blinder).

The interactive approach to the relative wage impact of unionism qualifies this conclusion somewhat. The proportion of black males who are union members is approximately the same as for whites, but the union relative wage effect is much greater for the former than the latter. Thus, the ratio of black to white wages was estimated by Ashenfelter to be 3.4 percent greater in 1967 than it would have been in the absence of unionism. On the other hand, the incidence of unionism among female workers is a third of its value for male workers, so unionism probably exacerbates the male-female wage differential. The incidence of unionism is greater in the lower than the higher occupational categories and so, by the Stafford estimates, is the relative wage effect, and this obviously tends to compress the wage structure. Thus, the conclusion that unions have little overall effect on income inequality is reinforced by the interactive approach.

III. Effect of Unionism on the Rate of Change of Money Wages

Like the other questions, this one has been around for a long time, and as in all but the last question, there remain many unresolved issues associated with it. Recent approaches to this problem have, of course, been in the context of the effect of unionism on the shape and position of the aggregate Phillips curve. First, reference to collective bargaining between unions and management was employed to justify variables, such as profit rates and the rate of price inflation, in some of the early wage equations (see G. L. Perry and O. Eckstein and T. Wilson). The significance of profits in the determination of union wages would appear to provide credence to bargaining models; it also seems to establish the functional distribution of income as an important factor in the inflationary process. However, more recent work has shown that, *whatever* the profit rate is doing on the right-hand side of wage equations, it has nothing to do with unionism. G. Pierson estimated separate wage equations for strongly unionized and weakly unionized industries and found that the profit rate was a more significant determinant of wages in the ***nonunion*** sector than in the union sector. Hamermesh (1970) constructed a wage series based on contract settlements and also found that profit rates did not have a significant effect on wage changes.

Second, there is a good deal of evidence that wages in the union sector are less sensitive to excess demand conditions than wages in the nonunion sector. The coefficient on the unemployment rate in the Hamermesh study was much smaller than in the earlier economywide equations. Ashenfelter, Johnson, and Pencavel estimated separate Phillips curves for unionized and nonunionized workers in manufacturing and found that the sum of the coefficients on the unemployment rate and its lagged value was $-.92$ in the nonunion sector and just $-.18$ in the union sector. Similar results were obtained by Hamermesh (1972). This finding is also consistent with the wage rigidity hypothesis advanced by Lewis (1963b) to explain variations in the union-nonunion wage differential over time.

Third, the importance of the question of the role of unions in the aggregate wage determination process rests on the degree to which nonunion wage changes are influenced by union wage changes and vice versa. E. Phelps has argued that since only one in four of all workers is a union member it is reasonable to ignore the influence of unionism. Similarly, R. Lucas and L. Rapping suggest that unions merely attempt to maintain a certain wage position relative to the nonunion sector and that it is therefore appropriate to analyze aggregate wage determination solely in terms of competitive factors. On the other hand, the presumption of models such as that of Eckstein and Wilson is that nonunion wages are largely influenced by union wages. Public policy over the past twelve years has reflected the latter view with the periodic imposition of voluntary or more-or-less mandatory controls on wage changes in the union sector. Which of these views is the more correct is an empirical question, but there has been little systematic work toward its resolution.

This completes my brief survey of recent trends in the economic analysis of trade unionism. My major conclusion is that while little progress has been made in modeling the behavior of trade unions—their objectives and the role of the bargaining process—there has been a great deal of sound empirical work on the economic effects of unionism. We know, for example, that unions have only a small effect on the distribution of income and that they tend to flatten the aggregate Phillips curve. We also know that the union-nonunion wage differential varies systematically across different parts of the labor force. There are many questions, some of which have fairly important policy implications, which have not been wrapped up, such as the effect of unionism on the position (as opposed to the shape) of the aggregate Phillips curve (or the "natural" unemployment rate). Perhaps in the next few years there will be an increase in the amount of resources devoted to a rigorous examination of these questions.

REFERENCES

A. A. Alchian and H. Demsetz, "Production, Information Costs and Economic Organization," *Amer. Econ. Rev.*, Dec. 1972, *62*, 777–95.

O. Ashenfelter, "Racial Discrimination and Trade Unionism," *J. Polit. Econ.*, May/June 1972, *80*, 435–64.

——— and G. E. Johnson, "Unionism, Relative Wages, and Labor Quality in U.S. Manufacturing Industries," *Int. Econ. Rev.*, Oct. 1972, *13*, 488–508.

———, ———, and J. H. Pencavel, "Trade Unions and the Rate of Change of Money Wages in United States Manufacturing Industry," *Rev. Econ. Stud.*, Jan. 1972, *34*, 27–54.

W. N. Atherton, *Theory of Union Bargaining Goals*, Princeton 1973.

A. S. Blinder, "Who Joins Unions?" Industrial Relations Section, mimeo., Princeton 1972.

M. J. Boskin, "Unions and Relative Real Wages," *Amer. Econ. Rev.*, June 1972, *62*, 466–72.

O. Eckstein and T. Wilson, "The Determination of Money Wages in American Industry," *Quart. J. Econ.*, Aug. 1962, *75*, 379–414.

D. Hamermesh, "Human Capital Theory—Its Importance to Labor Economics," *Ind. Rel. Res. Assoc. Proc.*, 1973, 225–32.

———, "Market Power and Wage Inflation," *Southern Econ. J.*, 1972, 204–12.

———, "Wage Bargains, Threshold Effects and the Phillips Curve," *Quart. J. Econ.*, Aug. 1970, *84*, 501–17.

G. E. Johnson and K. Youmans, "Union Relative Wage Effects by Age and Education," *Indus. and Lab. Rel. Rev.*, Jan. 1971, *24*, 171–79.

H. G. Johnson and P. Mieszkowski, "The Effects of Unionization on the Distribution of Income: A General Equilibrium Approach," *Quart. J. Econ.*, Nov. 1970, *84*, 539–61.

H. M. Levinson, "Unionism, Concentration, and Wage Changes: Toward a Unified Theory," *Indus. and Lab. Rel. Rev.*, Jan. 1967, *20*, 198–205.

H. G. Lewis, "Relative Employment Effects of Unionism," *Ind. Rel. Res. Assoc. Proc.*, 1963a, 104–15.

———, *Unionism and Relative Wages in the United States*, Chicago 1963b.

R. E. Lucas and L. A. Rapping, "Price Expectations and the Phillips Curve," *Amer. Econ. Rev.*, 1969, *59*, 342–50.

R. L. Oaxaca, "Estimation of Union/Nonunion Wage Differentials within Occupational/Regional Subgroups," Univ. of Massachusetts, mimeo., 1973.

J. H. Pencavel, "Wages, Specific Training, and Labor Turnover in U.S. Manufacturing Industries," *Int. Econ. Rev.*, Feb. 1972, *13*, 53–64.

G. L. Perry, "The Determinants of Wage-Rate Changes and the Inflation-Unemployment Tradeoff in the U.S.," *Rev. Econ. Stud.*, Oct. 1964, *31*, 287–308.

E. S. Phelps, "Money-Wage Dynamics and Labor-Market Equilibrium," *J. Pol. Econ.*, supp. 1968, *76*, 678–711.

G. Pierson, "Union Strength and the U.S. 'Phillips Curve,' " *Amer. Econ. Rev.*, June 1968, *58*, 456–67.

M. W. Reder, "Unions and Measurement: The Problems of Measurement," *J. Pol. Econ.*, Apr. 1965, *78*, 188–96.

A. Rees, *The Economics of Trade Unions*, Chicago 1963.

S. Rosen, "Trade Union Power, Threat Effects and the Extent of Organization," *Rev. Econ. Stud.*, 1969, *36*, 185–196.

S. Slichter, *Union Policies and Industrial Management*, Washington 1941.

F. P. Stafford, "Concentration and Labor Earnings: Comment," *Amer. Econ. Rev.*, Mar. 1968, *58*, 174–81.

A. Throop, "The Union-Nonunion Wage Differential and Cost-Push Inflation," *Amer. Econ. Rev.*, Mar. 1968, *58*, 79–99.

L. W. Weiss, "Concentration and Labor Earnings," *Amer. Econ. Rev.*, Mar. 1966, *56*, 96–117.

Economics of Time and Labor Supply

By H. G. Lewis*

The trail markers in recent research on labor supply are Clarence D. Long's book on the labor force; Jacob Mincer's paper on the labor force participation of married women; Gary S. Becker's paper on the allocation of time; William G. Bowen's and T. Aldrich Finegan's massive volume on labor force participation; the Becker-Gilbert R. Ghez monograph on the allocation of time and goods over the life cycle, which has been circulating in manuscript for several years but is not yet published; James Heckman's unpublished but forthcoming paper on shadow wages, market wages, and labor supply; the Glen G. Cain-Harold W. Watts volume which contains nine papers by a total of thirteen economists; and the Watts-Albert E. Rees *Final Report of the New Jersey Graduated Work Incentive Experiment.*

We now have literally hundreds of empirical labor force participation, weeks worked, hours worked per week and per year equations. What can they do for us in understanding time series variations in these various dimensions of labor supply? Differences in labor supply behavior among countries? Within countries by sex and other divisions of the population? The consequences of a national negative income tax scheme?

When belatedly I began the preparation of this paper, I became acutely aware that a convincing evaluation could not be accomplished single-handedly within a few weeks and in the scope of a short paper, even with the help of the recent Julie DaVanzo-Dennis N. DeTray-David H. Greenberg sensitivity analysis and Cain and Watts's effort at summary and synthesis. The sheer bulk of the pertinent literature and the seeming diversity of its findings are simply too great. Even now I have not read much of this literature, to say nothing of having studied, reflected upon, and reconciled it. I agreed reluctantly to prepare today's paper, but the assignment has aroused my curiosity so much that in the months ahead I am going to try to put the hours-of-work puzzle together. I am encouraged by an earlier, somewhat similar experience—that of evaluating research on the relative wage effects of unionism—to think that a substantial part of the puzzle can be assembled. What follows in this paper, therefore, is not a completed evaluation, but impressions, and these will be focused mainly on the negative income tax experiments.

One becomes most aware of the diversity of findings, of course, in a context such as that of predicting with some numerical precision the consequences of a negative income tax (*NIT*) scheme. For such a problem it can make a great deal of difference whether a relevant elasticity is 0.05 or 0.25. Yet it is important that we not lose sight of the forest because of the differences among the trees. We have come a long way toward understanding the division of labor within the family, not in family by family detail, but on the average, and this is no mean accomplishment. The findings here are so well known, I judge, that I won't try to summarize them. Furthermore, I have the impression that there is a considerable degree of agreement that long-run, uncompensated, own-wage elasticities of hours worked by married, prime-age males are negative and numerically closer to zero than to unity: a figure such as -0.15 is, I think, in the ballpark.

---

* Department of Economics, University of Chicago.

However, for the *NIT* problem we need to know more—the corresponding income elasticity, for example, and here my impression of the findings is one of marked diversity.

I turn now to the *NIT* experiments, especially the *Final Report* of the New Jersey-Pennsylvania experiment. First three general comments:

1. The *Final Report* almost defies anyone to study it. There is no detailed subject index, no list of tables, no detailed statistical appendix in which one can find answers to such questions as: What proportion of the white, continuous husband-wife, experimental families not on welfare received no *NIT* payments at all? One such payment? Two? And so on. Or what proportion of the male heads were multiple jobholders? I am pleased to hear that the *Final Report* is not the *last* report. I trust that the next *Final Report* will correct this defect of the first *Final Report.*

2. The four *NIT* experiments (New Jersey-Pennsylvania, Rural, Gary, and Seattle-Denver) are producing a unique set of panel data in which low-income families are presented with an alternative budget restraint involving much lower net wages and much higher nonwage income than their usual budget restraint. I hope that the data emerging from these experiments will be made generally available on tape or cards.

3. It should be surprising to some economists to learn that a final report on the first of the experiments already is available. The New Jersey-Pennsylvania experiment was designed, of course, to provide *NIT* payments to eligible families for only three years and the last such payment was made in September 1972. But presumably an *NIT* experiment has income effects and intertemporal substitution effects, not only during the experimental (or payment) period, but after it as well. I am disappointed that the experiment was not designed to continue the data assembly and analysis for a postexperimental period of two or three years, at least, and hope that some such effort will be made in one or more of the other experiments. Furthermore, the certain prospect this year of an *NIT* budget restraint option in the fairly near future will have income and substitution effects on this year's behavior of some families. Here, too, I am unhappy that some of the experimental families were not told at their enrollment in the experiment that, say, three years later they would become eligible for *NIT* payments for a period of three years, and that data were not gathered and analyzed for the preexperimental as well as the experimental period.

What did the New Jersey-Pennsylvania experiment discover? Although the counterintuitive findings for black families and, to a lesser extent, those for the Spanish-speaking families are a puzzle that deserves further study, I discuss here only the results for white families. First, some figures for the white, continuous husband-wife, experimental families just preceding their enrollment in the *NIT* experiment. Family size averaged close to six persons with an average of one to two children less than six years of age. Mean family hours of work per week was forty-two; mean family earnings $100 per week; mean family income $111 per week, with 23 percent of the families below the poverty income line, 34 percent at 100–124 percent of the line, and 43 percent at 125–150 percent of the line. The male heads averaged thirty-six years of age and ten years of schooling; 97 percent were labor force participants of whom 5 to 6 percent were unemployed; their mean hours of work per week was thirty-five and their mean wage about $2.50 per hour. Only one-sixth of the wives were labor force participants and the wives averaged about four hours of work per week and a wage of about $2.10.

The guarantees and implicit tax rates of the *NIT* schemes for these families varied from family to family; the mean guarantee was 90 percent of the poverty line and the mean tax rate was 51 percent. Most of these families resided in Scranton, Pennsylvania.

At first blush the low labor force participation rates for the wives and large family size may appear surprising. Recall, however, that only families with able-bodied male heads eighteen to fifty-eight years of age and with a 1968 family income not in excess of 150 percent of the poverty line were enrolled in the experiment and that the poverty line increased with family size.

Under a national *NIT* scheme of the kind examined in the experiment, many families will never choose the *NIT* option —that is, they will behave in such a way that they do not receive *NIT* payments at any time in their lifetime. We would like the experimental findings to tell us, first of all, the characteristics—in terms of family size, nonwage income, and wage profiles—of such families as they relate to the *NIT* parameters (guarantee, tax rate, poverty line). Second, for the families who would receive *NIT* payments, we would like to know their hours of work response to the *NIT*, both during the payment periods and at other times, stated as a function of the *NIT* parameters and the family's nonwage income, wage, and other relevant characteristics. That is, we would like the findings to come in the form of a *response function* that will tell us who will opt for the *NIT* and their responses.

The research staff of the experiment tried diligently to estimate such a function for responses during the three-year experimental period. For male heads of families with an average normal income (about 120 percent of the poverty line) under an *NIT* scheme with a guarantee of 100 percent of the poverty line and a 50 percent tax rate, they estimated that the effect of the *NIT* on mean hours of work per week was a reduction of about two and one-half hours during the middle eight quarters of the experiment and about seven and one-half hours during the last year. The disincentive effects, they estimated, were negligible for families with normal income near the break-even level of their *NIT* plan, but rose markedly as normal income declined below the break-even level. The estimated disincentive effects also rose substantially as the tax rate increased *or the guarantee fell*. The last result, of course, is surprising, demands explanation, and casts a cloud of uncertainty over the estimated response surface.

For the wives in these families, the labor force participation rate fell, because of the *NIT*, from about one-sixth to about one-twelfth and hours worked from about four per week to about two. The measured disincentives were greatest during the first two years, there was a general tendency for the disincentives to increase with the *NIT* guarantee, but variation in the *NIT* tax rate produced little variation in measured response. There is some evidence that the disincentive effects *increased* as normal income increased, a result that is not so surprising as at first it may seem.

Some of the derivatives of the estimated response functions with respect to guarantee, tax rate, and family income capacity are highly questionable and this is unfortunate. But put derivatives aside and simply evaluate the functions for intact white families with able-bodied male heads eighteen to fifty-eight years of age near the center of gravity of the experiment's observations (guarantee of 100 percent of the poverty line, tax rate of 50 percent, normal income about 120 percent of the poverty line). Are the estimated responses that I have cited what we can expect for such families under a national and permanent *NIT* with the same parameters during periods in which they are receiving payments? I see no reason for supposing that these estimates are grossly misleading for a national, permanent *NIT* during its initial years for such families that expect to receive the *NIT* payments only temporarily—that is, for a relatively small fraction of their remaining working years. I would expect the disincentives to become somewhat larger after the *NIT* had become more familiar and better understood and especially to be larger for families expecting to receive *NIT* payments more or less permanently.

For illustrative purposes, consider the following lifetime utility function and budget restraint for a husband(m)-wife(f) family:

Utility function:

$$u = \left(\sum_{t=1}^{n} u_t^{(\sigma-1)/\sigma} \right)^{\sigma/(\sigma-1)},$$

where σ is either zero or 0.2;

$$u_t = \text{Min}\,(u_{mt}, u_{ft}); \qquad 5u_{mt} = (\lambda_{mt}^{-1} + x_{mt}^{-1})^{-1};$$
$$u_{ft} = (9\lambda_{ft}^{-1} + x_{ft}^{-1})^{-1}$$

Budget restraint:

$$A + \sum_{t=1}^{n} 100(w_{mt} + w_{ft}) = \sum_{t=1}^{n} \pi_t u_t;$$

where u denotes utility, t the week in a remaining n-week lifetime, 100 is the assumed number of hours per week disposable for either leisure time λ or market work, x is market goods consumption, the w's are real market wage rates per hour in terms of market goods, A is initial net real assets, π_t is the shadow price of u_t, and I assume for simplicity a zero interest rate. The illustration is a simplification, to put it mildly, of the Becker-Ghez model and the substitution elasticities in the illustration are not those estimated by them, though they are not, I think, preposterously different.

Assume that A is zero and that w_m and w_f do not vary over the lifetime, with $w_m = \$2.56$ per hour and $w_f = \$1.64$ per hour. At these values the husband works 40.4 hours per week, the wife 16.8 hours, and together they earn $130.91 per week. I take these as the *base* values for measuring the impact of the *NIT*.

Now offer the family a permanent *NIT* option with a 50 percent tax rate and a weekly guarantee of $67.50 (a weekly break-even of $135.00). Turn to Table 1, row 1. The first six columns show the ratio of hours worked under this *NIT* experiment to base hours. The husband's hours fall by 30 percent and the wife's by 62 percent.

Suppose that the preceding *NIT* option is offered now for a period amounting to one-fifth the expected remaining lifetime of the family. Then the lifetime income effects of the *NIT* are smaller than on line (1), but there may be intertemporal substitution effects as well. There are no intertemporal substitution effects on line (1). Line (2) assumes that the intertemporal substitution elasticity σ is zero, so that line (2) differs from line (1) during the experimental period only in having smaller income effects. Hence during the experimental period, hours worked fall by less on line (2) than on line (1). Notice that hours worked fall slightly after the experimental period because of the increase in lifetime income produced by the *NIT*. Line (3) assumes an intertemporal substitution elasticity equal to 0.2. The intertemporal substitution possibilities present on line (3) have two consequences. First, the *income* effects are larger than on line (2) and these will tend to reduce hours both during and after the experiment. Second, and more important, intertemporal substitution between the experimental and later periods by itself causes hours worked to fall during

TABLE 1—ILLUSTRATION OF DISINCENTIVE EFFECTS OF *NIT* EXPERIMENTS IN A LIFETIME CONTEXT

Experimental Period and Value of σ	Ratio of Husband's Hours to Base Hours (Base=40.4)			Ratio of Wife's Hours to Base Hours (Base=16.8)		
	During Experiment (1)	After Experiment (2)	Lifetime Mean (3)	During Experiment (4)	After Experiment (5)	Lifetime Mean (6)
(1) Lifetime; σ irrelevant	0.70	—	0.70	0.38	—	0.38
(2) $\frac{1}{8}$ of lifetime; $\sigma=0$	0.76	0.99	0.95	0.55	0.98	0.89
(3) $\frac{1}{8}$ of lifetime; $\sigma=0.2$	0.62	1.01	0.93	0.12	1.04	0.85
(4) $\frac{1}{8}$ of lifetime; $\sigma=0^{a}$	0.99^{b}	0.996	0.995	0.56	0.99	0.90
(5) $\frac{1}{8}$ of lifetime; $\sigma=0.2^{a}$	0.99^{b}	1.004	1.00	0.44	1.01	0.90

[a] See text for explanation.

[b] By assumption; see text.

the experimental period and to *rise* thereafter. Thus, in the illustration, hours worked fall by more on line (3) than on lines (1) and (2) during the experimental period, but rise slightly after the experiment. From the comparison of lines (2) and (3) with line (1), it appears to be unclear whether a short-duration *NIT* experiment will lead to under- rather than overestimation of the disincentive effects of a permanent *NIT* scheme.

However, lines (1), (2), and (3) assume that workers freely can choose their hours of work at wage rates given by the market —that is, that employers are indifferent with respect to the hours of work of their employees. Thus, for example, on line (3) the husband, in response to the *NIT*, cuts his hours per week from 40.4 to 25.0 during the experiment and then raises his hours to 40.9 thereafter, presumably without having to change employers. It seems more reasonable to suppose that changes in hours of work of this magnitude sometimes, perhaps often, would involve changing jobs with associated search costs and, for some, loss of seniority (specific human capital) benefits. (In this connection it would be useful to know what fraction of the male heads in the experiment were multiple jobholders, their hours worked on the secondary job, and the impact of the *NIT* experiment on these hours.) It also seems likely that such costs typically will be greater for male heads with a more or less permanent, full-time attachment to the labor force than for the wives, who usually have less firm labor force attachments.

Lines (4) and (5) of the illustration assume that these job-change costs are sufficiently great for the husband in the context of the short-duration experiment that he avoids changing jobs, though he does cut his hours slightly during the experiment. (This assumption is consistent with the finding in the experiment that the experimental *NIT* did not lead to an increase in job turnover for male heads, though it does not explain reduced job turnover.) In other respects the model for lines (4) and (5) is the same as that for the preceding lines. In particular, it is still assumed that the wife freely chooses her hours.

The effect of the restraint on the husband's hours assumed in the illustration is to reduce both the lifetime income effects of the experiment, as can be seen by comparing line (2) with line (4) for the wife's hours, and especially the intertemporal substitution effects, as shown by a similar comparison of line (3) with line (5).

In the illustration the wife cuts her hours during the experiment by less on lines (4) and (5) than on line (1). I don't mean to argue that this proves that the short duration of the experiment led to underestimation for married women of the disincentive effects of a similar but permanent *NIT* program, but only that the *Final Report* conceivably may not be correct in arguing the opposite. Of course, I have set the model up in such a way as to get the results on lines (4) and (5), but that does not make the results wrong.

As for the husbands, there is another reason for suspecting that the long-run disincentive effects may have been underestimated, namely, that substitution possibilities within the household between time and market goods may be substantially larger in the long run than in the short run. Do-it-yourself that is neither thumb twiddling to the doer nor malingering to the observer requires planning and preparation.

REFERENCES

G. S. Becker, "A Theory of the Allocation of Time," *Econ. J.*, Sept. 1965, *75*, 493–517.

——— and G. R. Ghez, "The Allocation of Time and Goods over The Life Cycle," unpublished, Apr. 1972.

W. G. Bowen and T. A. Finegan, *The Economics of Labor Force Participation*, Princeton 1967.

G. G. Cain and H. W. Watts, *Income Maintenance and Labor Supply*, Chicago 1973.

J. DaVanzo, D. N. DeTray, and D. H. Greenberg, *Estimating Labor Supply Response: A Sensitivity Analysis*, Santa Monica 1973.

J. Heckman, "Shadow Prices, Market Wages, and Labor Supply," unpublished, Nov. 1972.

H. G. Lewis, "Hours of Work and Hours of Leisure," *Annual Proceedings of the Ind. Rel. Res. Assn.*, 1956, 196–206.

C. D. Long, *The Labor Force Under Changing Income and Employment*, Princeton 1958.

J. Mincer, "Labor Force Participation of Married Women," in H. G. Lewis, ed., *Aspects of Labor Economics*, Princeton 1963.

H. W. Watts and A. E. Rees, *Final Report of the New Jersey Graduated Work Incentive Experiment*, Madison 1974.

DISCUSSION

PAULA E. STEPHAN, Georgia State University: Gregg Lewis focuses upon possible effects of the negative income tax (*NIT*) within the traditional framework of the work-leisure choice. Lewis applies the familiar tools of comparative statics, albeit to multiple time periods. Lewis' results are of interest since they remind us that the *NIT* market-work disincentive results are dependent upon the length of time of the *NIT* and the intertemporal elasticity of substitution. With regard to the first of these points, we should be cautioned that the *NIT* work effects are dependent upon the length of time of the *NIT* relative to the horizon of the individual, not relative to his working life. Indeed, one likely reason the poor are poor is because of an abbreviated horizon.

The *NIT* is a useful vehicle for analyzing aspects of time allocation. I find it somewhat disappointing, however, that Lewis chooses to analyze the question within the parochial framework of utility maximization. It is useful to point out that other approaches exist. In particular, Lewis does not mention that the *NIT* can affect human capital accumulation. Consider the human capital approach which states that the objective of the individual is to maximize the present value of income net of investment costs. Within this framework, it can be shown that by lowering the after-tax value of the individual's time the *NIT* will lower the indirect cost of human capital investment as well as the present value of the benefit stream arising from the human capital investment. Because the *NIT* does not alter the cost of direct inputs used in human capital production, however, the *NIT* will cause the present value of benefits arising from the human capital investment to decrease relative to the cost of investment. As a result, the *NIT* leads to a flatter human capital profile. One consequence of this may be to keep persons permanently on the negative income tax roles. Since earnings are lower in the early and late years of the life cycle, the *NIT* may expect to enroll primarily the young and the old. If, however, the *NIT* has human capital disincentive effects, today's young enrollees may prove to be eligible in the mid-years as well as in the late years of their lives.

Just as the Lewis work disincentive effects depend upon the length of time of the *NIT* relative to the horizon of the individual, so too do the human capital disincentive effects. While the present value of benefits arising from human capital production depends upon the relative length of time of the *NIT*, the costs do not. As a result, a nonpermanent *NIT* will have smaller human capital disincentive effects than a permanent *NIT*. Indeed, when the *NIT* is not permanent, it is possible to conceive of a situation where the *NIT* operates to augment the quantity of human capital produced by the individual. If, for example, the *NIT* is relatively short-lived, the cost of human capital production may well decrease more than the resulting benefits and we may find persons accumulating human capital inventories.

The Lewis view sees the wage rate as exogenous and hours of work as endogenous and demonstrates that the *NIT* can lower the number of hours worked. The human capital view sees the combined number of hours of work and training as exogenous and the wage rate as endogenous and demonstrates that a permanent *NIT* can flatten the wage profile. Both are, of course, partial views of the world. Yet these partial views are seldom joined. As a profession we teach our students that the hours-of-leisure variable is solved in a utility maximization framework while the wage-rate variable is solved by the technique of lifetime income maximization.

Recent work by G. S. Becker and G. R. Ghez attempts to unite these two approaches by depicting the life cycle as a series of static utility-maximizing decisions superimposed upon an endogenous wage rate which changes over time as the human capital stock is altered by the individual's investment decision. Following the Becker-Ghez approach, one could argue that the *NIT* acts to augment leisure not only because of utility considerations, but also because human capital considerations lead to a dampening of the value of time.

The problem with the Becker-Ghez approach is that partial solutions derived statically do not necessarily provide the solution to a dynamic problem. Because time can be allocated between labor, leisure, and training, human capital decisions are affected by utility considerations as well as profitability considerations. Further, human capital decisions as well as utility considerations affect hours of leisure. Consequently time allocation over the life cycle should be treated in the context of a dynamic model which recognizes that today's decisions influence the decisions to be made tomorrow. Such a model was presented by Frank Stafford and me at the 1972 meetings in Toronto. Further developments along these lines have been set forth in a paper by A. S. Blinder and L. W. Weiss. The results of these models are of interest because one can deduce properties of the time paths of labor, leisure, and training over the life cycle. Unfortunately, as is the case with most dynamic models, it is not possible to obtain explicit solutions to these models. Yet it does seem that insights into possible effects of the *NIT* on life-cycle labor supply and human capital accumulation could be achieved through simulating some of the functional forms of the dynamic models developed to date.

Apart from the fact that dynamic models may be successful in explaining life-cycle labor supply and earnings, the conclusion we are forced to draw is one which is far too familiar to economists: namely, a tradeoff exists among the progression of models we have described. As we endogenate the variables, insights provided in the form of explicit solutions vanish. On the other hand, this process is useful since it will at least alert us to the interdependencies that exist and to how these interdependencies may alter the conclusions of simpler models. As a result, this process will allow us to test the conclusions of our simpler models. Properties of these models which prove to be unaltered by the more complete models can then be considered robust.

PAST DEVELOPMENTS AND FUTURE PROSPECTS FOR ETHNIC MINORITY GROUPS

Three Phases of Cliometric Research On Slavery and its Aftermath

By ROBERT WILLIAM FOGEL*

The cliometric investigation of slavery and its aftermath is now nearly two decades old. Initiated by the path-breaking paper presented by Alfred H. Conrad and John R. Meyer in 1957, it has been carried forward in more than 100 subsequent papers, doctoral dissertations, and books.[1] This extensive scholarly effort has contributed not only to the reshaping of the discipline of economic history but has greatly altered our understanding of the nature of slave society and, most important, has helped to reveal the remarkable record of black achievement in the protracted and difficult struggle to overcome racist oppression. It is not possible in this brief paper to review adequately the many contributions to the reinterpretation of the slave economy. I would, however, like to outline what I believe to be the three main phases of cliometric research on the subject and to suggest some priorities for future research.

During the first phase, which extended roughly from 1957 to 1969, cliometric research was focused on three issues: the profitability of an investment in slaves to slaveholders; the economic viability of the slave economy; and the effect of slavery on southern economic growth. All three of these issues were, of course, raised in the original Conrad and Meyer paper, but not all of them were explored to the same extent by either Conrad and Meyer or by their successors. Much of the initial work focused on the issues of profitability and viability, partly, perhaps, because these seemed to be more tractable than the question of growth, partly because the evidence bearing on profitability appeared to be more readily available.

The debates on profitability have involved both theoretical and empirical issues. Much of the discussion has turned on the equation originally employed by Conrad and Meyer to estimate the rate of return on male slaves. They became involved in the difficult task of estimating the average amount of land and capital employed per slave, although this issue could have been finessed by employing a slightly different variant of their basic equation.[2] More vexing were the implicit assumptions that all twenty-year-old slaves lived for exactly thirty additional years (the approximate life expectation of male slaves at age twenty) and that between ages twenty and fifty average annual earnings of adult males were invariant. Robert Evans, Jr. and N. G. Butlin improved on the demographic assumption by providing estimates of the age-specific death rates of slaves that were quite similar, even though Butlin's estimates were derived from the Jamaican life-table and Evans' were derived from the published data in the *U.S.* census of 1850. It was not, however, possible to obtain estimates of the age-earnings profile (the average earnings of a cohort of slaves at each age from birth to the death of the last member of the cohort) from published sources.

---

* Professor of economics and history at the Universities of Chicago and Rochester. Research on the paper was supported by a grant from the National Science Foundation.

[1] See, in addition to the references at the end of this paper, those in Fogel and S. L. Engerman 1974a. Convenient collections of cliometric papers on slavery are H. G. J. Aitken, Engerman and E. D. Genovese, and Genovese and R. N. Rosett.

[2] See E. Saraydar, R. Sutch, and J. D. Foust and D. E. Swan. Also, see Fogel and Engerman 1974a, vol. 2, p. 66.

The principal contribution on the issue of viability prior to 1969 was made by Y. Yasuba, who pointed out that the rate of return on the purchase price of a slave by itself was insufficient to establish that slavery was economically viable. Yasuba proposed to measure viability by the capitalized rent on slaves, that is, by the difference between the market price of a slave at a given age and the cost of rearing him to that age. Yasuba's approach was further refined by T. Bergstrom who proposed to use the capitalized rent at age zero (the zero-age price of a slave), which has been called the price of a "birthright."[3]

The principal contribution on the question of southern economic growth was made by R. A. Easterlin who, after working out regional income accounts, discovered that between 1840 and 1860 the per capita income of the South grew by 1.3 percent per annum. Engerman, by taking account of Texas, raised this figure to 1.7 percent, thus revealing that during two decades preceding the Civil War the slave South was growing more rapidly than the free North. While these findings clearly refuted the belief that the antebellum South was stagnating, they did not settle the counterfactual questions of whether slavery retarded economic growth during the antebellum era or whether the legacy of slavery retarded growth during the postbellum era.[4]

The principal feature of the second phase is the massive effort to uncover primary data bearing on the operation of the slave economy. This phase of the cliometric investigation of slavery was launched in the early 1960's when Parker and Robert E. Gallman began the collection of a random sample from the manuscript schedules of the *U.S.* census for 1860. By 1968 they had not only retrieved, but put in machine-readable form, information on 5,000 southern farms and 40,000 slaves. Shortly thereafter they, various of their students, and others began to illuminate the ongoing debates on slavery by providing far better evidence than had hitherto been available. Even more important, this new data base stimulated a substantial enlargement of the set of issues under consideration. Wright (1969) examined the distribution of wealth among free whites in the South and probed into the question of economies of scale in cotton production; Foust analyzed the role played by the yeoman farmer in the westward expansion of cotton production; Gallman, as well as R. C. Battalio and J. Kagel, estimated the degree to which large slave plantations were self-sufficient in the production of food.

The search for primary data was given new impetus by the attempt to use the geometric index of total factor productivity to compare the relative efficiency of northern and southern agriculture for the year 1860. The year 1860 was singled out because the availability of the Parker-Gallman sample permitted a comparison of the efficiency of large and small farms within the South. When, however, the first crude measurements produced the unexpected result that southern slave agriculture was substantially more efficient than northern free agriculture, it became clear that the evidence contained in the Parker-Gallman sample, by itself, was inadequate for dealing with this complex issue.

---

[3] Engerman and I independently hit upon the same approach in 1969.

[4] Papers probing these questions include W. N. Parker, C. Pope, P. Passell and G. Wright, Wright 1974b, and L. J. Kotlikoff and S. Pinera.

In our 1971 essay, Engerman and I argued that what showed up as the superior efficiency of slave agriculture in these indexes was probably due, in large part, to the mismeasurement of various of the inputs, particularly the labor input. For these indexes had been constructed on the assumption that hours worked per man-year were the same on southern slave plantations as on the free farms of the North. This, we conjectured, was probably wrong; it was likely that slaves worked more hours per day and more days per year than free farmers. We therefore launched a search for the additional data needed to measure the length of the slave and free work-years, to take adequate account of the much higher proportion of women and children in the slave than in the free labor force, and to determine the skill composition of the respective labor forces.

This search—a collective effort that was carried forward by a team of over thirty researchers, assisted and advised by more than a dozen leading scholars of antebellum history—brought forth an enormous increase in the available body of evidence bearing on slavery. One component of the new data base, a sample of 5,700 estates containing information on 77,000 slaves who lived between 1770 and 1865, was retrieved from the probate records of southern courts. This body of information made it possible to construct age-price profiles for slaves, from which the long-missing age-earnings profile could be computed. The age-earnings profile was not only needed for the resolution of the issue of profitability, but was critical for the issue of efficiency as well, since it provided the basis for converting the unadjusted measures of the labor input of women, children, the aged, and the disabled into full-hand equivalents. About 10 percent of the estates in the probate sample also provided information on the skill composition of the slave labor force.

A second component of the new data base is a sample of the business records of roughly 100 large plantations containing either highly detailed information on the organization of production (including records of the daily activities of each slave in the labor force) or demographic information needed to adjust the labor input of women. For many of these plantations the demographic record extends over a period of at least 75 years, thus bearing on the lives of three or more generations of slaves. This source, which contains information on roughly 20,000 slaves, also yielded much important evidence on the skill composition of the slave labor force. Still other important bodies of systematic data bearing on the issue of efficiency were found in slave-trading and mortality records. The various new samples, together with the Parker-Gallman sample, contain information on over 250,000 slaves, or about 4 percent of the slaves who ever lived in the United States or the colonies out of which it was formed.[5]

Surprisingly, the evidence culled from the new sources did not sustain our original conjecture that the relative efficiency of slave agriculture was due largely to the failure to measure the labor input in man-hours. Quite the contrary, the evidence indicated that the work-year of free northern farmers averaged over 3,100 hours while that of slaves on large plantations averaged about 2,800 hours (Olson and Fogel).

---

[5] More detailed descriptions of the types of information contained in the probate and large-plantation samples are contained in J. Metzer, R. H. Steckel, and J. F. Olson and Fogel. The other samples are described in Fogel and Engerman 1974a and 1974b.

If the finding that slave work-year was actually 10 percent shorter than that of free farmers in the North seems too ridiculous to be taken seriously, it is because most of the traditional portrayals of slavery have been based on the naive assumption, to which we originally subscribed, that the greater exploitation of slaves must have taken the form of the lengthening of the work-year.

But as every trade unionist who has grappled with the problem of speedup knows, it is possible to increase the real length of the work-year, even while reducing the number of clock-time hours, by increasing the intensity of work per hour. This point, which was emphasized in *Time On The Cross*, has been further illuminated by the recent work of Metzer and Olson. Their preliminary findings, summarized in the paper by Olson and Fogel, indicate that slaves worked 82 percent more intensely per hour than did the free farmers of the North. Thus the typical slave on a large plantation produced as much in thirty-three minutes as did the typical free farmer in an hour. It follows that if one measures the length of the work-year not in clock-time hours but in hours of equal labor intensity (equal-efficiency hours), the slave work-year was actually 65 percent longer than the free work-year.

Once it is recognized that the fundamental form of the exploitation of slave labor was through increasing intensity per hour rather than through an increase in clock-time hours per year, many paradoxes resolve themselves. The longer rest breaks during the work day and the greater time off on Sundays for slaves than for free men appear not as boons that slaveowners granted to their chattel, but as conditions for achieving the desired level of intensity. The finding that slaves earned slightly more income per clock-time hour loses its startling quality when it is realized that their pay per equal-efficiency hour was 37 percent less than that of free men.

No aspect of *Time On The Cross* has occasioned more debate than the finding that slaves were more efficient and harder working than their free counterparts. Some scholars have suggested that the total-factor-productivity index used to compare the relative efficiency of free countries cannot be applied to a comparison of the free North and the slave South—that in this instance a morally weighted index of efficiency is required.[6] However, the issue of the relative efficiency of slave labor did not originate with *Time On The Cross*. It is an issue with a long history that traces back to such commentators on slavery as Adam Smith, Alexis de Tocqueville, Cassius Marcellus Clay, Hinton Rowan Helper, Frederick Law Olmsted, and John E. Cairnes. Is the geometric index of total factor productivity appropriate to the resolution of the question of the efficiency of slave labor as that question actually evolved in historical literature?

Much of chapters five and six of *Time On The Cross* is devoted to reviewing both the pre- and post-Civil War debates on the inefficiency of slave agriculture and slave labor. From this review it is clear that what the critics of slavery meant was: other inputs held constant, but substituting slave for free labor, and slave managers for free managers, the output of slave farms would be much less than the output of free farms. About this "fact" Clay, Helper, Olmsted and most other antislavery critics had no doubt. This confidence stemmed from the conviction that slavery "degraded" labor, that slavery turned plantation owners into "idlers," that, "comparing man with man," slave laborers were less than half as productive as whites, that Africans were "far less adapted for steady, uninterrupted labor than we are," and that "white labourers of equal intelligence and under equal stimulus will cut twice as much wood, split twice as many rails, and hoe a third more corn in a day than Negroes." (See Clay, p. 204, and Olmsted, pp. 91, 467–68.)

---

[6] See P. David and P. Temin 1974, especially pp. 778–83.

Nor does the geometric index of total factor productivity do violence to the issue of efficiency as it was perceived and discussed by the principal scholars who preceded us. Certainly neither U. B. Phillips, nor L. C. Gray, nor R. B. Flanders, nor R. R. Russel, nor K. M. Stampp were talking about a morally weighted measure of productivity, but of the comparative efficiency of slave and free labor, in just the manner that this issue was raised by the antebellum critics of slavery. In Gray's words:

> The discussion of efficiency of Negro slave labor has been befogged by a failure to distinguish between those industrial qualities peculiar to Negro labor and those attributable to the influence of slavery. There has been a tendency to assume that all the laziness and incapacity exhibited by the Negro are ascribable to his status as a slave. This confusion has been largely the outgrowth of the arbitrary assumption that slave labor must be inherently less economic than free labor because slave labor is given grudgingly and half-heartedly, while free labor is spurred by self-interest. Turgot and Adam Smith developed this attitude, and it soon became a commonplace of economic generalization, accepted even by some Southern writers. Francis A. Walker gave the formula an eloquent presentation: "The whip," he wrote, "only stimulates the flesh on which it is laid. It does not reach the parts of the man where lie the springs of action. No brutality of rule can evoke even the whole physical power of a human being." Cairnes recognized three fundamental elements of inefficiency in slave labor: "It is given reluctantly; it is unskillful; it is wanting in versatility.". . .
>
> Many contemporary writers, including some of proslavery sympathies, assumed that slave labor was intrinsically uneconomical as compared with free labor in general farming. Yet, in the physical operations of general farming slaves frequently displayed considerable skill. A contemporary writer asserted, "In the use of the axe, hoe, scythe and cradle, or driving and training teams, the practiced Negro of Georgia is surpassed by few white men I ever saw." It was admitted by some observers that slaves could cultivate as large an acreage of corn or wheat per hand as could a free laborer. Even in the care of livestock it was possible to find reliable individual Negroes to whom these functions could be entrusted. The inferiority of Negro slave labor in general farming was not due to the incapacity of the Negro to perform the manual operations involved, most of which, in that period when farm machinery was but little employed, were not more complex than in producing the staples. Indeed, general farming with slave labor at times proved successful. Where it was unprofitable, this was due mainly to three conditions: (1) A larger amount of supervision per unit of labor was required than in less diversified forms of agriculture, which is also true when free Negro or white labor is employed; (2) In a number of regions general farming was severely handicapped for lack of a home market; (3) The competitive superiority of cotton and sugar enabled planters employing virgin soils to offer higher prices for slaves than could be afforded in general farming. [pp. 462–63, 469]

The index of total factor productivity is, then, an appropriate instrument for resolving the issue of efficiency as it has been raised and debated in the historical literature for more than a century. And the results of the measurements performed with this index directly refute the calumny that blacks were less efficient or less hardworking than their white counterparts. Not only has this old view been refuted, but the weight of evidence now indicates that blacks on large plantations produced nearly twice as much output per unit of time as whites. Surely this is the wrong time to assert that output per manhour, standardized for other inputs, is a morally biased and unacceptable measure of the efficiency of labor.

Of course, the fact that blacks were more efficient under slavery than were free laborers does not imply that black workers are inherently superior to white workers. There is every reason to believe that if whites had been forced to labor under the gang system and blacks had been free farmers, then white slaves would have been more efficient than free blacks. It was the system that forced men to work at the pace of an assembly line (called the gang) that made labor more efficient, not race. Nor does greater efficiency mean greater good. Indeed the whole point of Table 3 of *Time On The Cross* was to demonstrate that freedom has value and that the loss of freedom was greater than the gain in measured output (vol. 1, pp. 244–46).

The principal feature of the third phase of cliometric research on slavery is the shift of emphasis from how the slave system worked to the recovery of black history. The onset of the third phase has not brought the second to an end. Rather, the two phases coexist, each giving vitality to the other. The effort to direct cliometric research toward the recovery of black history is, in fact, a direct consequence of the massive hunt for evidence bearing on the relative efficiency of slave agriculture. In the course of the search for the data needed to refine the measures of labor input, scholars have uncovered evidence that reveals new aspects of the widespread and successful struggle of blacks to overcome the worst features of the slave system.

The theme of black struggle under slavery is, of course, hardly new. But until recently many scholars have tended to treat this issue as if the main efforts of slaves to promote their self-interests were either revolutionary forms of violence, quasi-revolutionary flight for freedom, or covert but politically self-conscious forms of opposition that have been called "day to day resistance."[7] Deliberate warriors against oppression have, however, except perhaps for historically fleeting moments, always been an extremely small fraction of the population from which they sprang. What of the mass of slaves? Were they, as was so long argued, largely broken and corrupted by the system? Was there no room within the system, short of overt or covert political warfare, for those who had to bear the yoke of oppression to struggle for their self-interest? And if there was, what forms did these struggles take and with what success did they meet?

We have not yet acquired the hard evidence needed to answer these questions adequately, although there is considerable ground for believing that the desired evidence is recoverable. But what is already in hand clearly undermines the view that blacks were typically so crushed by the system of slavery that they were unable to advance their self-interests, to thwart the forces that militated against family solidarity, to acquire skills, and to become effective workers.[8] The findings on total factor productivity, for example, have exposed the frail basis for the contention that slaves were typically uneducable and incompetent laborers who "got into very bad habits of doing as little as possible" (Conrad et al., p. 539). Interestingly, some scholars have responded to the evidence indicating that blacks were actually harder working and more efficient than their white counterparts either by suggesting reverse racism or by arguing that for blacks to have been diligent implies that they must have accepted the morality and objectives of the master class (David and Temin

---

[7] See Stampp, pp. 108–09; also, Fogel and Engerman 1974a, vol. 2, pp. 208–10 and 224–25.

[8] See S. C. Crawford 1975; also, S. Stuckey, G. P. Rawick, R. S. Bryce-LaPorte, John W. Blassingame, and Genovese.

1975). This view of life under slavery surely oversimplifies the complex and contradictory forces that influenced the behavior of both slaves and masters and underestimates the extent to which the acquisition of skills and the improvement in the material conditions of life were wrung from the system.

Whatever the motives of the master class, slaves had good reasons of their own to be diligent and to acquire skills. Available evidence indicates that the level of the material conditions that slaves could obtain for themselves and for their children, the capacity to resist encroachments on the integrity of their families, and the opportunity to enlarge the scope for freedom of choice were all related to their effectiveness as workers (cf. Crawford 1975). On the matter of the diet, for example, statistical analysis of the Fisk University and WPA narratives of ex-slaves reveals that slaves who relied purely on the rations provided by masters had less varied diets than those who cultivated their own plots or engaged in hunting and fishing.[9] Analysis of the notarial records in New Orleans reveals that slave artisans and craftsmen were less likely to have been sold than ordinary field hands (Olson).

It is also a gross oversimplification to suggest that skills acquired by slaves were merely those that their masters conferred on them. While it may be true that economic forces generated by the market tended to make it profitable for masters to promote skill formation among their slaves (cf. C. D. Goldin), it also appears that powerful political forces operated in the opposite direction. It is now more than half a century since C. G. Woodson chronicled the efforts of blacks to acquire education during the antebellum era, finding that the effort reached its maximum extent and success prior to 1825. It is hard to come by the data needed to test Woodson's conjecture that between this date and 1860 black literacy rates declined by half under the impact of the reactionary political assault that made it a crime to teach slaves to read and write. On the other hand, there is strong evidence that between 1850 and 1860 this assault reduced by nearly half the number of large plantations on which blacks were the chief nonownership managers.[10] Increased pressure to enforce laws requiring the presence of white overseers appears to have been the principal reason for this decline.

The finding that in certain respects the attack on the material conditions of black life may have been more acute during the postbellum than during the antebellum era has led to a substantial increase in cliometric research into southern and black history covering the whole period from 1865 to the present. J. D. Reid, Jr., S. DeCanio, R. Ransom and Sutch, R. Shlomowitz, Wright, and R. Higgs are examining the institutional arrangements upon which the new economic order was constructed, seeking the mechanisms that undermined black economic life. Richard B. Freeman is examining the growth of discrimination against blacks in the labor market from 1870 through the depression of the 1930's, as well as assessing the factors which explain the reversal of this trend, especially after 1964. M. Hopkins and N. S. Cardell are measuring the change in the real income of blacks between 1850 and 1880. Engerman, E. Meeker, J. E. Eblen, and R. Farley are studying trends in black nuptiality, fertility, and mortality and their relationship to changing economic and social conditions.

One of the lessons of the latest phase of cliometric research is that the investigation of black economic history is an exercise in political economy in the fullest sense of the term. This is even more true for the postbellum era than for the antebellum era. The work of Reid, DeCanio, and Freeman strongly suggests that much of the deterioration in the material conditions of black life that took place after 1870 was due less to the operation of the unfettered forces of the market than to the use of state power to transfer income and wealth from blacks to whites. The evidence marshaled by Freeman indicates that licensure laws were used to squeeze blacks out of skilled crafts in which they had been strongly represented, while educational restrictions were used to prevent them from entering new crafts. He has also found that during the depression of the 1930's the number of blacks employed by state and local governments in the South fell by close to half even though the total number of government jobs in most southern states increased by the same fraction. These states, Freeman concludes, "appear to have replaced blacks with whites in menial jobs that had been undesirable in good times but became valuable in bad times" (ch. 8, p. 19).

---

[9] Fogel and Engerman 1974c. The problems involved in the quantification of the data in the ex-slave narratives and pitfalls of inferences made from these data are discussed in Crawford 1974.

[10] See Genovese and Rosett, especially the debate on the morning of October 24 and the response to the discussion on October 26.

* * *

In closing this essay it is worth recalling that during the 1960's there were raging debates as to whether black studies and black history had a legitimate place in the university curriculum. Many feared, not entirely without reason, that they would be artificial creations or, at best, "soft" subjects that undermined scholarly standards. How ironic it is, then, that black history has become the leading arena in the effort to apply rigorous quantitative methods to historical research.

REFERENCES

H. G. J. Aitken, *Did Slavery Pay?*, Boston 1971.

R. C. Battalio and J. Kagel, "The Structure of Antebellum Southern Agriculture: South Carolina, A Case Study," *Ag. Hist.*, Jan. 1970, *44*, 25–37.

T. Bergstrom, "On the Existence and Optimality of Competitive Equilibrium for a Slave Economy," *R. of Econ. Stud.*, Jan. 1971, *38*, 23–36.

J. W. Blassingame, *The Slave Community*, New York 1972.

R. S. Bryce-LaPorte, "Slaves as Inmates, Slaves as Men: A Sociological Discussion of Elkins' Thesis," in A. J. Lane, *The Debate Over Slavery*, Urbana 1971, 269–92.

N. G. Butlin, *Ante-bellum Slavery—Critique of a Debate*, Canberra 1971.

C. M. Clay in H. Greeley, ed., *The Writings of Cassius Marcellus Clay: Including Speeches and Addresses*, New York 1848.

A. H. Conrad and J. R. Meyer, "The Economics of Slavery in the Ante Bellum South," *J. Pol. Econ.*, Apr. 1958, *66*, 95–130.

S. C. Crawford, "Toward a Quantitative Analysis of the Data Contained in the WPA and Fisk University Narratives of Ex-slaves: Some Preliminary Findings," Report 7475-1, Workshop in Econ. Hist., Univ. of Chicago, 1974.

———, "Autobiography of an Oppressed Class: A Quantitative Study of the WPA and Fisk University Narratives of Ex-slaves," Univ. of Chicago, doctoral dissertation, expected 1975.

P. David and P. Temin, "Slavery: The Progressive Institution?," *J. Econ. Hist.*, Sept. 1974, *34*, 739–83.

——— and ———, "Capitalist Masters, Bourgeois Slaves," *J. Interdisc. Hist.*, Winter 1975, *7*.

S. DeCanio, *Agriculture in the Postbellum South: The Economics of Production and Supply*, Cambridge, Mass. 1974.

R. A. Easterlin, "Regional Income Trends, 1840–1950," in S. Harris, ed., *American Economic History*, New York 1961, 525–47.

J. E. Eblen, "New Estimates of the Vital Rates of the United States Black Population during the Nineteenth Century," *Demography*, May 1974, *11*, 301–19.

S. L. Engerman, "Changes in Black Fertility, 1880–1940," Univ. of Rochester, mimeo., 1974.

——— and E. D. Genovese, *Race and Slavery in the Western Hemisphere: Quantitative Studies,* Princeton 1974.

R. Evans, Jr., "The Economics of American Negro Slavery," in Universities-Nat. Bur. Comm. for Econ. Res., *Aspects of Labor Economics,* Princeton 1962.

R. Farley, *Growth of the Black Population: A Study of Demographic Trends,* Chicago 1970.

R. B. Flanders, *Plantation Slavery in Georgia,* Chapel Hill 1933.

R. W. Fogel and S. L. Engerman, "The Relative Efficiency of Slavery: A Comparison of Northern and Southern Agriculture in 1860," *Explorations in Econ. Hist.,* Spring 1971, *8,* 353–67.

——— and ———, *Time on the Cross,* 2 vols., Boston 1974a.

——— and ———, eds., "Evidence Relevant to the Post-Publication Debate on *Time on the Cross,*" Univ. of Rochester, mimeo. 1974b.

——— and ———, "Further Evidence on the Nutritional Adequacy of the Slave Diet," Univ. of Rochester, mimeo. 1974c.

J. D. Foust, "The Yeoman Farmer and Westward Expansion of U. S. Cotton Production," Univ. of No. Carolina, doctoral dissertation 1967.

——— and D. E. Swan, "Productivity and Profitability of Antebellum Slave Labor: A Micro-Approach," *Ag. Hist.,* Jan. 1970, *44,* 39–62.

R. B. Freeman, "The *U.S.* Discriminatory System: The Economics of Discrimination against Black Americans, 1870–1970," unpublished manuscript.

R. E. Gallman, "Self-Sufficiency in the Cotton Economy of the Antebellum South," *Ag. Hist.,* Jan. 1970, *44,* 5–23.

E. D. Genovese, *Roll, Jordan, Roll: Afro-American Slaves in the Making of the Modern World,* New York 1974.

——— and R. N. Rosett, eds., "Proceedings of the MSSB-University of Rochester Conference on *Time on the Cross,*" Oct. 24–26, 1974.

C. D. Goldin, "The Economics of Urban Slavery: 1820 to 1860," Univ. of Chicago, doctoral dissertation 1972.

L. C. Gray, *History of Agriculture in the Southern United States to 1860,* 2 vols., Washington 1933.

H. R. Helper, *The Impending Crisis of the South: How to Meet It,* Cambridge 1968.

R. Higgs, "Race, Tenure, and Resource Allocation in Southern Agriculture, 1910," *J. Econ. Hist.,* Mar. 1973, *33,* 149–69.

M. Hopkins and N. S. Cardell, "The Cost of Freedom: A Comparison of the Real Income of Free Blacks in 1880 with the Real Income of Slaves in 1850," in progress.

L. J. Kotlikoff and S. Pinera, "The Old South's Stake in the Interregional Movement of Slaves, 1850–1860," Harvard, mimeo. 1974.

E. Meeker, "Mortality Trends of Southern Blacks, 1850–1910: Some Preliminary Findings," *Explorations in Econ. Hist.,* forthcoming.

J. Metzer, "Efficient Operation and Economies of Scale in Ante-Bellum Southern Plantations," Report 55, Dept. of Econ., Hebrew Univ., 1974.

F. L. Olmsted, *The Cotton Kingdom,* New York 1953.

J. F. Olson, "The Occupational Distribution and Characteristics of Skilled Slaves during the Late Antebellum Era," Univ. of Rochester, doctoral dissertation, expected 1975.

——— and R. W. Fogel, "Clock Time vs. Real Time: A Comparison of the Lengths of the Work Years in Northern and Southern Agriculture," Univ. of Rochester, mimeo. 1974.

W. N. Parker, "Slavery and Southern Economic Development: An Hypothesis and Some Evidence," *Ag. Hist.,* 1970, *44,* 115–25.

P. Passell and G. Wright, "The Effects of Pre-Civil War Territorial Expansion on the Price of Slaves," *J. Pol. Econ.,* Nov./Dec. 1972, *80,* 1188–1202.

U. B. Phillips, *American Negro Slavery: A Survey of the Supply, Employment and Control of Negro Labor As Determined by the Plantation Regime,* New York 1918.

C. Pope, "The Impact of the Ante-Bellum Tariff on Income Distribution," *Exploration* *in Econ. Hist.,* Summer 1972, *9,* 375–421.

G. P. Rawick, *From Sundown to Sunup: The Making of the Black Community,* Westport 1972.

J. D. Reid, Jr., "Aspects of Sharecropping," Univ. of Chicago, doctoral dissertation 1974.

R. R. Russel, "The General Effects of Slavery upon Southern Economic Progress," *J. of Southern Hist.,* Feb. 1938, *4,* 34–54.

E. Saraydar, "A Note on the Profitability of Ante Bellum Slavery," *Southern Econ. J.,* Apr. 1964, *30,* 325–32.

R. Shlomowitz, "Postbellum Southern Labor Contracts, 1865–1868," Report 7475-11, Workshop in Econ. Hist., Univ. of Chicago, 1974.

K. M. Stampp, *The Peculiar Institution: Slavery in the Ante-Bellum South,* New York 1956.

R. H. Steckel, "Slave Marriage, Fertility, and Society," Report 7374–4, Workshop in Econ. Hist., Univ. of Chicago, 1973.

R. Sutch, "The Profitability of Ante Bellum Slavery—Revisited," *Southern Econ. J.,* Apr. 1965, *31,* 365–77.

——— and R. Ransom, "What Was Freedom Worth?" unpublished manuscript 1974.

S. Stuckey, "Through the Prism of Folklore: The Black Ethos in Slavery," *Massachusetts Rev.,* Summer 1968, 417–37.

C. G. Woodson, *The Education of the Negro Prior to 1861: A History of the Education of the Colored People of the United States from the Beginning of Slavery to the Civil War,* Washington 1919.

G. Wright, "The Economics of Cotton in the Antebellum South," Yale Univ., doctoral dissertation 1969.

———, "Cotton Competition and the Post-Bellum Recovery of the American South," *J. of Econ. Hist.,* Sept. 1974a, *34,* 610–35.

———, "The Economic Analysis of Time on the Cross," paper presented to the MSSB-Univ. of Rochester Conf. on *Time on the Cross,* Oct. 24–26, 1974b.

Y. Yasuba, "The Profitability and Viability of Plantation Slavery in the United States," *The Econ. Studies Quart.,* Sept. 1961, *12,* 60–67.

Public Policy and Black Economic Progress: A Review of the Evidence

By BERNARD E. ANDERSON AND PHYLLIS A. WALLACE*

The 1960 decade was characterized by a marked improvement in the economic progress of the black community, narrowly defined as gains in black income and employment. The decade also brought new initiatives in public policy designed to influence economic growth and stability, human resource development, and equal employment opportunity. Although few of the economic and social programs developed during the 1960's were focused primarily on blacks, it is likely that the wide array of public policies pursued during the decade influenced both the pace and direction of changes in black employment and income.

The purpose of this paper is to summarize and evaluate the evidence compiled by economists to explain the pattern and determinants of change in the employment and income of blacks during the past decade. In the past few years, economists have given much attention to issues related to racial inequality, with much of the research conducted by a new generation of young economists who grew to maturity in the environment of the civil rights revolution when new perceptions and values on race relations were formed. Powerful analytic tools of analysis, conceptual paradigms, and data sources have been explored in an effort to better understand and explain the persistence of racial inequality. Yet there remains a considerable diversity of opinion on the meaning and significance of black progress in employment and income, and the extent to which the observed progress can be attributed to different types of public policy. It may be useful to summarize very briefly the broad trends in black income and employment between 1960 and 1974, and then evaluate the views of economists regarding the role of public policy in reducing racial inequality.

I. Trends in Black Employment and Income

During the 1960's, blacks experienced significant gains in the labor market, but the progress was uneven and discontinuous. Between 1961 and 1969, the number of blacks in the civilian labor force grew by 1,151,000, with adult black women accounting for 57 percent of that number. Black employment grew by slightly more than 1.5 million, distributed between 601,000 adult black men, and 755,000 adult black women. Black unemployment fell by 400,000, concentrated among black men, and the black unemployment rate declined sharply from 12.4 to 6.4 percent. Most of these gains were registered during the years 1965 through 1969, but the pause in economic activity between 1969 and 1971 reversed the trend in an unfavorable direction. During that two-year period blacks bore a disproportionately large share of the increase in unemployment, but failed to recoup their loss during the mild economic expansion of 1972 and 1973.

The 1960 decade also witnessed a marked degree of occupational upgrading within the black labor force. The proportion of all blacks in white-collar jobs rose from 16.1

---

* The Wharton School, University of Pennsylvania, and Sloan School of Management, Massachusetts Institute of Technology, respectively.

percent to 27.9 percent, with marked growth in the number of blacks in professional, technical, and clerical occupations. The proportion of blacks employed as skilled craftsmen also increased markedly, while those in unskilled labor and domestic service declined sharply. Blacks registered respectable gains in some industries such as banking, insurance, and utilities where black workers in the past had not been greeted with open arms. Most of the occupational gains were concentrated among younger blacks who entered the labor market with relatively greater schooling and job skills in comparison with older blacks in the labor market.

Yet despite the favorable occupational gains, blacks were still less likely than whites to get into the higher paying jobs, and were still concentrated in low-wage industries. Moreover, blacks continue to comprise a disproportionately large share of the net addition of labor in occupations and industries with limited opportunities for upgrading and with high sensitivity to cyclical fluctuations in the economy.

The ambivalence of black economic progress may be observed in earnings trends. The median income of black families doubled between 1960 and 1970, while the black/white income ratio increased from .55 to .64. The earnings improvement was evident across a wide spectrum, with the percent of black males earning less than $3,000 declining markedly from 40 percent in 1959 to 22 percent in 1969. Among black females, the proportion earning less than $3,000 per year declined from 75 to 52 percent. Conversely, the percent of black men earning more than $10,000 rose from 3 to 9 percent, while the proportion of black women in that earnings category increased from less than 1 percent to 5 percent.

Still, the dollar gap between whites and blacks continued to rise throughout the decade, reaching close to $4,700 in 1972. Also, while the number of blacks in poverty declined by 2.2 million between 1959 and 1970, the flight from poverty among blacks occurred at less than half the rate experienced by low-income whites. Further evidence of disparity is found in the black/white differential in subemployment, i.e., the combined proportion of those who are either unemployed, discouraged workers, involuntary part-time workers, or low-paid heads of families. Despite the economic gains of the past decade, almost 30 percent of the blacks in urban areas were subemployed in 1972, compared with less than 10 percent of all whites.

One of the most disturbing trends during the past decade was the continued deterioration in the labor market experience of black youth. Black teenagers experienced rates of unemployment of 25 percent and above throughout the decade, and withdrew from the labor force in alarming numbers. The dramatic decline in labor force participation among black youth greatly exceeded the experience of young whites and approached a magnitude difficult to explain in terms of the rise in school attendance among black teenagers.

II. Interpretation of the Trends

A survey of the economic literature reveals a general consensus that the relative status of blacks, especially the more educated persons, increased significantly during the past decade. Moreover, studies by most of the researchers confirm the concentration of gains among those who entered the labor market after 1965 (see F. Welch, W. Vroman, R. Freeman, R. Hall and R. Kasten, D. Bell, and J. Gwartney). A variety of data sources have been examined in the research including the decennial census, Current Population Survey, Continuous Work History Sample of the Social Security Administration, the 1967 Survey of Economic Opportunity, and the National Longitudinal Surveys from the

Parnes project at Ohio State University. Some of these data were not available for research before the 1960's.

Although the economic researchers agreed that the relative economic status of blacks had improved, they articulated competing hypotheses on the major reasons for the significant increase in the black/white income differential that had been relatively constant for much of the pre–1960 period. Economists also differed in their views on the source of continued black/white disparities in income and employment. A high level of economic activity, particularly in the latter part of the 1960's, the improved quality of education for blacks, and the reduction in labor market discrimination were identified as the major determinants. Since earnings gains were larger than what might have been expected from tight labor markets, the debate centered on the relative effects of education compared with the implementation of fair employment laws in generating the observed trends.

One optimist hailed the "collapse in traditional discriminatory patterns" and argued that the increase in black incomes during the post–1964 period was influenced primarily by the antidiscrimination efforts of the federal government (Freeman). On the other hand, O. Ashenfelter and J. Heckman, analyzing data on 40,445 establishments reporting to the U.S. Equal Employment Opportunity Commission, found that black male employment increased by only 3.3 percent more in firms with government contracts than in those with no contracts between 1966 and 1970. This difference, statistically significant at the 1 percent level, suggests that antidiscrimination activity exerted much less impact on black employment gains than estimated by some researchers (Freeman, Hall and Kasten, Welch).

At least three secondary issues raised by these studies deserve a brief comment. In cross-sectional and trend analyses several researchers found evidence that the relative income gains for blacks in the South improved markedly after 1966. In a study of black employment in the South, Ray Marshall and a team of investigators found that the South's traditional institutions including institutionalized discrimination are being eroded by forces associated with industrialization and urbanization. However, the relative gains made by blacks in the South did not alter the fundamental imbalance between black and white employment patterns. Even after the noticeable progress, blacks were two to three times as likely as whites to be clustered in low-paying operative, laborer, and service occupations (Marshall and V. Christian).

Other researchers (S. Masters, A. Van Adams and G. Nestel, L. Weiss and J. G. Williamson) were concerned about the role of migration in improving the economic position of blacks. They argued that black migrants to metropolitan communities in the North had higher incomes and less unemployment than blacks born there. "Young blacks reared and educated in the urban ghettoes of the non-South seem likely to pay a considerable premium for this environment in terms of subsequent earnings" (Van Adams and Nestel).

Educational advancement among blacks seemed to generate significant gains in employment and earnings, but the evidence does not suggest that education alone holds the key to economic equality. Although black income increased progressively with education, the marginal gain in income associated with additional increments of education appeared somewhat less for blacks than for whites. Moreover, with the black/white income ratio for males falling with increased education, the potential for improving the relative economic position of black men through increased schooling was somewhat limited. Over the decade, it appears that increased educational attainment favored black women more than black men (Freeman, Hall and Kasten, Vroman).

Another area of interest pertaining to human capital considerations is the impact of manpower programs on black economic progress. Blacks were very prominent participants in the phalanx of manpower programs initiated during the 1960's, and accounted for about 46 percent of the 8.9 million program enrollees recorded between 1965 and 1972. Blacks, however, were disproportionately concentrated in the programs which emphasized objectives other than the acquisition of marketable job skills. Perhaps as a result, the limited data on economic impact by race generally indicate that manpower programs had a very small effect on the relative employment and earnings position of blacks (C. Perry et al.). Further, the limited gains generated by the programs seem more the result of improvements in participant employment stability than in greater access to higher paying jobs.

In order to place the record of gains and contributing forces in clearer perspective, the following comments may be useful: (1) Despite past increases in black occupational status and earnings, large disparities still remain. Such gaps are evident throughout the labor force, but are especially acute among black youth and among prime age black males, whose labor force participation has declined at an alarming rate. (2) The increase in the relative economic status of black women can be considered dramatic only in relation to the restrictions placed on the upward mobility of white female workers. As indicated by Patricia Gurin, a social psychologist, "the normative stance in social science in which black men are compared to white men and black women to white women has contributed to the view that black women are somehow better off than everyone except white males." This bias obfuscates the reality of the black woman's economic position which is at the bottom of the income pyramid. (3) Increased investment in the education of blacks is a necessary but not sufficient condition for reducing the black/white income differential. Institutionalized discrimination is both subtle and pervasive and affects access to jobs, training, and promotion opportunities, thereby constraining the maximization of the black income-earning potential.

In regard to human capital investment policies, it would not be correct to conclude that manpower training programs have no role to play in increasing the relative status of blacks. Instead, it would be more accurate to say that the programs most likely to benefit blacks are those which emphasize the acquisition of marketable job skills. Even these programs, however, are unlikely to be effective if they are not linked to antidiscrimination activities. (4) Aggregative policies designed to stimulate full employment and tight labor markets have been necessary for black economic progress, but have been insufficient to produce economic equality. Some blacks register significant gains in an environment of strong labor demand while many others are left behind in a residue of unemployment, underemployment, and poverty. Among those who seem least affected by tight labor markets are black youth and inner city residents (P. Wallace).

At this point it might be appropriate to speculate on the potential effects of the current economic problems on black income and employment. Because black employment deteriorates more rapidly than that of white workers during an economic decline, it is quite likely that blacks may experience a reversal of the relative economic gains achieved in recent years. Economists are now predicting a rise in unemployment that will exceed the level attained at any time during the postwar period. If both the depth and duration of the current recession come close to the consensus projections, the fragile gains of blacks during the 1960's may be lost.

Such a pessimistic prognosis may be warranted for the following reasons. First, a deep recession will markedly worsen the position of lower-income blacks who typically receive an even smaller proportion of total money income than do lower-income white families in periods of reduced economic growth. Second, the position of black youth may well become intolerable as even larger numbers withdraw from the labor force in response to the inability to find jobs. This development will further reduce black family income and truncate the human capital improvement that occurs when young people receive on-the-job training. Third, a deep recession will, by its very nature, influence a wide spectrum of industries and occupations and will adversely affect employees with many years of seniority. For this reason, the current economic conditions may be more threatening to black progress than the relatively mild recessions of the past two decades when the average unemployment rate remained below 6.0 percent. In the current environment, it would not be unlikely to observe occupational downgrading and increased competition across education groups that may adversely affect black employment in the lower entry level white-collar jobs[1] where blacks made significant gains since 1965.

Finally, rising unemployment and the resulting increase in job competition may affect the pursuit of affirmative action and equal job opportunity. As the recession deepens, the conflict between equal opportunity and job security will generate great pressures on the personnel practices of firms in a wide range of industries. Even though antidiscrimination activity will undoubtedly continue, the direction and intensity of such efforts will certainly be affected by the political realities imposed by worsening economic conditions.

In conclusion, economists should be encouraged to continue their examination of the relative economic status of blacks in large part because economic research may provide support for urgently needed policies for moderating the degree of income inequality. It is also necessary to undertake additional research on the status and progress of other racial and ethnic minorities in the work force. Despite the lack of consensus on the relative influence of different types of public policy, it is clear that a variety of policies must be pursued simultaneously. Continued improvements in the quantity and quality of the type of economic research initiated during the 1960's on racial inequality will contribute significantly to a clearer notion of the proper mix and linkage of public policies required to achieve economic parity for racial and ethnic minorities.

REFERENCES

O. Ashenfelter and J. Heckman, "Measuring the Effect of an Antidiscrimination Program," *Evaluating the Labor Market Effects of Social Programs*, Industrial Relations Section, Princeton Univ., work. pap. no. 52.

D. Bell, "The Economic Impact of the Social Legislation of the 1960's on Blacks in the Labor Market," Rand Corp., p-5202.

A. Brimmer, "Widening Horizons: Prospects for Black Employment," May 5, 1974.

R. Freeman, "Changes in the Labor Market for Black Americans, 1948–72," *Brookings Papers on Economic Activity*, 1973, *1*, 67–120.

C. Gilroy, "Black and White Unemployment: The Dynamics of the Differential," *Monthly Lab. Rev.*, Feb. 1974, *97*, 38–47.

P. Gurin and C. Gaylord, "Sex-Role Constraints: The College Educated Black Woman," Univ. of Michigan, unpublished paper 1974.

J. Gwartney, "Employment Discrimination, Productivity Factors and Income Differentials between Whites and Nonwhites," *Amer. Econ. Rev.*, June 1970, *60*, 396–408.

———, "Changes in the Nonwhite/White Income Ratio, 1939–67," *Amer. Econ. Rev.*, Dec. 1970, *60,* 872–883.

R. Hall and R. Kasten, "The Relative Occupational Success of Blacks and Whites," *Brookings Papers on Economic Activity,* 1973, *7,* 781–797.

A. Kohen, *Determinants of Early Labor Market Success Among Young Men: Race, Ability, Quantity and Quality of Schooling,* Center for Human Resource Res., Ohio State Univ., Jan. 1973.

R. Marshall, "The Economics of Racial Discrimination," *J. of Econ. Lit.,* Sept. 1974, *12,* 849–871.

——— and V. Christian, eds., *The Employment of Southern Blacks,* Salt Lake City forthcoming.

S. Masters, "Are Black Migrants from the South to the Northern Cities Worse Off than Blacks Already There?" *J. of Human Resources, 7,* 411–423.

C. Perry, B. Anderson, R. Rowan, and H. Northrup, *The Impact of Government Manpower Programs,* Philadelphia 1975.

E. Phelps, "Economic Policy and Unemployment in the 1960's," *The Pub. Int.,* Winter 1974, 30–46.

L. Thurow, *Poverty and Discrimination,* Washington 1969.

A. Van Adams and G. Nestel, *Interregional Migration, Education and Poverty in the Urban Ghetto: Another Look At Black-White Earnings Differentials,* Center For Human Resources Res., Ohio State Univ., Sept. 1973.

U.S. Department of Labor, *Manpower Report of the President, 1974,* Washington 1974.

W. Vroman, "Changes in Black Workers' Relative Earnings: Evidence from the 1960's," in G. M. Von Furstenberg et al., eds., *Patterns of Racial Discrimination,* Lexington 1974.

P. Wallace, *Pathways to Work,* Lexington 1974.

F. Welch, "Educational and Racial Discrimination," in O. Ashenfelter and A. Rees, eds., *Discrimination in Labor Markets,* Princeton 1973.

———, "Black-White Differences in Returns to Schooling," *Amer. Econ. Rev.,* Dec. 1973, *63,* 893–907.

L. Weiss and J. G. Williamson, "Black Education, Earnings and Interregional Migration: Some New Evidence," *Amer. Econ. Rev.,* June 1972, *62,* 372–83.

Factors Affecting the Future Economic Prospects of Minorities

By DAVID H. SWINTON*

This paper discusses the general characteristics of a market economy that determines the future prospects of minorities. We are principally interested in the long-run implications of automatic market processes for the economic position of minorities. We will give only brief consideration to the role of nonmarket factors. The principal conclusion of our analysis is that the market processes will not lead to automatic improvement in the situation for minority groups. We further conclude that the likely future scenario is for slight deterioration in the minorities' relative position if the market process is left alone. And finally, we conclude that sociopolitical factors can be expected to intervene into the market process with undetermined results.

The overall structure of economic well-being in a private property market economy is determined by the technical and organizational characteristics of productive activities and the stock of both human and nonhuman resources. For any given structure of productive activities and resources, we can postulate the existence of an equilibrium distribution of economic well-being. The market mechanism should automatically bring about the equilibrium distribution. Our discussion concentrates on the factors determining the distribution of this equilibrium well-being between majority and minority groups.

Most economic theorists would probably accept the proposition that a market economy could reach an equilibrium without full employment and, in fact, without even guaranteeing survival of all of its members. Moreover, it is also the case that our modern economy is characterized by differentiated opportunity structures. The existence of unemployment and differentiated opportunity structures creates the possibilities for relative differences in individual economic well-being in equilibrium.

Since capitalistic societies are characterized by the selfish pursuit of individual self-interest, individuals can be expected to utilize whatever techniques are available to attain the higher valued opportunities. Thus, individuals have an incentive to enter coalitions to influence the personal allocation of opportunities, where a basis for stable coalitions exists and where the benefits from coalition membership are greater than the cost. In complex, multiracial and capitalistic societies, racial or ethnic coalitions arise to influence the allocation of benefits whenever clear group identifications exist and the potential power of one group is greater than the potential power of the other.

Thus, at any point in time, the relative and absolute position of the minority will be determined by three principal factors: the level and structure of opportunity in the society, the distribution of resources, and the extent of discrimination. The level and structure of opportunity determine what is available in the given society. The distribution of resources determines an individual's technical suitability or eligibility for the available opportunities. And discrimination is a part of the process which allocates the scarce opportunities.

---

* State University of New York, Stony Brook.

Changes in the position of the minority over time result from changes in these basic factors. We shall discuss each of these factors in turn while holding the others constant.

The degree of inequality in the structure of opportunities and the quantity of opportunities relative to the population are the first factor. If there are wide disparities between the rewards of various opportunities, the minority will be relatively worse off. On the other hand, in structures characterized by narrow differences in rewards, the minority-majority differences will be less. Likewise, where there are few opportunities relative to the population, the minority will suffer a relatively high level of unemployment. Expansion of opportunities at rates faster than the rate of growth of resource populations will cause expanded employment of minority-owned resources. Failure of opportunities to expand as fast as populations will cause relative and absolute unemployment of minority resources. Increases in the range and variance of value will lower the relative value of minority-owned resources.

The significant question is whether the economy tends to create, through the automatic workings of the market, the factors which would lead to relative improvement in the minority position. Any automatic tendencies to cause convergence in differentiated reward structures or to create full employment of resources will improve the minorities' relative welfare over time.

Although full employment was frequently assumed in the past, the economic processes which would bring it about were never clear, not even in the theoretical perfectly competitive economy, if the least bit of realism was introduced. The wage rate that would bring about full employment may not be above subsistence and, in fact, may not even be nonnegative. Moreover, in models where one recognizes the relative fixity of capital/labor ratios and the limitations of effective demand, the likelihood of a less than full employment equilibrium at any given level of resource and income is apparent. In the real world there are also numerous barriers to resource mobility that create relatively inflexible resource prices, especially in the downwards direction. This inflexibility is relatively immune to the existence of unemployed resources. Yet it is the flexibility of resource prices that is at the heart of the market's automatic adjustment capability.

There may actually be a tendency to unemployment and underemployment of resources given the investment propensities of capitalists. Automation, for instance, clearly must tend to increase the unemployment of human resources, other things remaining equal. The greater use of large-scale techniques with large initial requirements for capital also clearly lowers the value and increases the unemployment of resources employed in small-scale activities. Increasingly, important resource shortages and limitations of effective demand may also tend to increase unemployment of resources. Increased international competition and the investment propensities of multinational corporations also limit the employment of domestic resources. The possibility of increasing resource unemployment seems almost unavoidable when we combine these demand side changes with current sociological trends, which are tending to increase the relative supply of human resources.

In a competitive economy, rates of return to equivalent resources in different employments would tend to equality over time through the workings of the arbitrage processes. This would cause convergence in the net value of employed minority and majority resources. However, deviations from perfect competition that interfere with resource mobility prevent the arbitrage processes from operating effectively. Given such imperfections, which certainly are characteristic of the American economy, the market cannot automatically eliminate value differences. Thus, such differences can be expected to continue over time if the market is left to its own ways.

The second major factor influencing the relative well-being of minorities is the structure of resource ownership. The larger the proportion of resources owned or controlled by the minority, the better off is the minority. As before, our principal interest is in the implications of market operations for the dynamic path of the resource distribution.

If there were no extra market intervention, there would be little reason to expect greater equality in the racial distribution of resources. The accumulation of resources depends upon savings, and since savings is a function of current income, which in turn depends upon current resources ownership, it is clear that those who currently own the most resources will do the most savings.

The distribution of human capital over time is probably not as directly tied to the initial distribution because human capital in general cannot be inherited. Moreover, there is a large-scale public role in the creation and distribution of human capital through the public school system. Nonetheless, there is still a strong tie between initial wealth and the ability to attain human capital since much human capital, especially high quality types, must be purchased through the private market. Even within the sphere of public supply, there are quantitative and qualitative differentials in human capital, and the access to the better qualities here also depends upon initial wealth. Much human capital is acquired in conjunction with the pursuit of various opportunities, so that the distribution of this type of capital is subject to the same considerations as the distribution of opportunities.

However, over time we cannot expect the slight equalizing tendency in human capital to be sufficient to cause significant reduction in current disparities in overall resource ownership. The recent trends respecting racial differentials in the quality of human capital acquired in the public schools would indicate that part of the quantitative gains has been offset by qualitative deterioration. The public's willingness to expand the public role sufficiently to eliminate both the quantitative and qualitative differences which remain appears limited.

We must conclude, therefore, that there are no automatic forces operating in the market to eliminate resource differences over time. In fact, the automatic tendency of the market would be to increase differences in resource ownership. This tendency is partially offset by public activity especially in the provision of human capital development. However, the overall impact of this activity is limited and in any event it is not automatic. Barring additional intervention, the market will not significantly reduce the disparities in resource ownership. This factor, therefore, cannot lead to improvement in the well-being of minorities over time.

The final factor influencing the economic future of minorities is the extent of discrimination. The impact of this factor on the well-being of minorities at any point depends upon the intensity of discrimination. The more intense the discrimination, the worse off is the minority. The intensity of discrimination will depend upon the cost and benefits of discrimination. The greater the benefit, the greater the discrimination, and the greater the cost, the smaller is the level of discrimination.

At any point in time, the benefits will depend upon the level and structure of opportunity and the racial distribution of resources. The greater is the quantity of resources held by the minority, the more the benefit from discrimination. Similarly, the more structural inequality and the lower the level of opportunities, the more benefit there is from discrimination. The cost also depends fundamentally on the distribution of resources and the level and structure of opportunity. Here the greater the amount of resources held by the minority, the higher the cost of discrimination. However, the lower the level of opportunities and the greater the structural inequality, the less costly is discrimination.

Here again we are also interested in the dynamic impact of automatic market forces. To reduce discrimination, these forces would have to reduce the benefits of discrimination relative to the cost. However, both the benefits and the cost of discrimination are connected to the market through the distribution of resources and the structural characteristics of production. Therefore, the economic incentive to discriminate can only be reduced over time by changing these underlying factors. We have already seen that there are no identifiable automatic forces working toward greater equality in the distribution of resources or less differentiation in the structure of rewards. These factors, therefore, cannot be counted on to lower discrimination.

However, there could well be more inequality in the distribution of resources. This would lower the cost as well as the benefit from discrimination. Only if the benefit is lowered more than the cost will overt discrimination lessen. However, this will not lead to improvement in the situation of minority groups because the lower discrimination will be offset by lower relative resources.

The dynamic paths taken by the economy during this century have worked in two directions at once. Technological change in southern agriculture has tended to increase the relative unemployment of minority resources which were heavily engaged in low-valued agricultural pursuits at the turn of the century. However, the rate of expansion of opportunities in the nonagricultural sectors was sufficient to absorb a large proportion of the minority resources that could not be employed in agriculture. This not only prevented the unemployment of minority resources, but also increased the absolute and overall relative value of minority-owned resources. However, these migration processes have led to deteriorating relative value of minority-owned resources in regions outside of the south since 1950. Moreover, these intersectoral and interregional arbitrage possibilities have practically vanished, and the rate of expansion of opportunities in the society as a whole has not been sufficient to absorb all of the resources in the past couple of decades. As a result, minorities find themselves increasingly unemployed, underemployed and outside of the labor force.

Moreover, throughout this period the overall inequality in the value of opportunities has not significantly narrowed. And during the past twenty-five or so years, the degree of inequality in the structure of opportunities has probably increased. This fact alone should be expected to lead to relative increases in the gap between the value of majority and minority resources.

We can therefore draw the overall conclusion that automatic market adjustment processes will not tend to improve the relative situation of the minority in the future. The three determining factors mentioned *could* lead to improvement, stability, or deterioration since all three types of changes would be consistent with market equilibrium and would be produced under the appropriate circumstances. There is no automatic tendency for the economy to produce either result. In the current situation, however, the dynamic implications of the market process for the structure of opportunities and the distribution of resources imply at best slight deterioration in the relative well-being of minorities in the future. Thus, if the selfish maximization of individual welfare continues to be the prevailing behavioral norm, the free market, left to itself, will attain an economic equilibrium path which preserves the relative minority-majority welfare disparity.

This economic equilibrium, however, will not necessarily be consistent with overall sociopolitical equilibrium. The continued existence of relative disparities could lead to the intervention of noneconomic factors into the equilibrium determination processes as a result of efforts by the minority to reduce the disparities. Thus, efforts would be made either to alter the three determining factors or to alter the distribution-determining processes, i.e., change the economic system.

Noneconomic action can be carried out either by public authorities or private individuals and coalitions. In the latter case, it is the minority that will act since it is the minority that will be dissatisfied with the free market solution. The success of minority efforts will depend upon their ability to reduce the benefits or raise the cost accruing to the majority as a result of the current equilibrium. The idea is to introduce noneconomic cost-raising factors to the degree where it becomes optimum for the majority to alter the status quo solution. Such alteration involves reducing discrimination, redistributing resources or altering the structure of opportunities.

It is not clear that there exists a set of noneconomic actions available to the minority that will move the status quo equilibrium as far as desired. It is probably the case, if parties are rational, that there is a set of actions that will move the status quo equilibrium in the desired direction. Obviously, however, noneconomic actions involve cost of some kind. Moreover, the majority may be able to respond with action short of altering the status quo equilibrium. However, this action itself must involve cost so there must be a negotiating range. Whether or not there is movement in the desired direction will depend upon the relative strengths of the two parties. One should also note that it is possible that such tactics could lead to a lower status quo equilibrium for the minority. However, pursuit of these points is beyond the scope of this paper.

Moreover, it is also not clear that conflict situations such as the above can be treated with the limited concept of rationality implicit in what we have done. *Discrimination practiced over long periods of time develops its own dynamics.* Thus, even though the narrowly defined economic cost of discrimination exceeds the economic benefit of discrimination, the process may continue for some period of time, either because of expectations of future benefits, failure to perceive the changed situation, or the existence of noneconomic benefits. Failure to perceive the loss in net benefits will be particularly likely when, as is the case with discrimination in the United States, the benefits and cost are distributed differently. Moreover, the public authorities may absorb most of the cost of maintaining the status quo, further hiding the cost from individuals. Thus, individuals will not associate the benefits with the cost, or will believe that the cost can be reduced without reducing the benefits.

Rational public authorities might also be induced to intervene in the market equilibrium process if the disparity prevents the attainment of social equilibrium. The intervention by the public authority will be brought about by its responsibility to maintain the social equilibrium where deviations from social equilibrium either threaten the social stability or the social welfare. The inducement results from the fact that minorities will favor alteration of the status quo equilibriums along racial lines and majorities will prefer maintenance of status quo equilibriums. This results in conflict as discussed above. The public authority must undertake activities that reduce the conflict. This can be accomplished in at least two ways. The first way is to suppress the minority cost-raising activity. The second is to alter the status quo equilibrium to reduce the majority-minority disparity. This latter action may give rise to majority cost-raising activities.

The action taken by the public authority will depend upon the objectives of the public authority, which in turn follows from the processes by which objectives are reached and the relative weight of minorities in these sociopolitical processes. This, however, is far beyond the scope of this paper but it is clear that we can by no means assume that public action will be favorable to the minority.

In any case, in view of the fact that the market will not automatically cause improvement, any chance of assuring improvement in future prospects for minorities lies in some kind of sociopolitical intervention in the market processes. The phenomenon of discrimination is itself a social intervention into the determination of market equilibrium. There is nothing sacred about the market equilibrium which itself is just the outcome of a particular set of sociopolitical processes. Therefore, additional intervention can certainly be justified whenever the outcome is undesirable. In view of the difficulties inherent in changing status quo solutions, it should be apparent that the minority must be willing to make sacrifices and persevere in order to improve its future prospects.

DISCUSSION

Ronald L. Trosper, University of Washington: My mandate here is to extend the three papers to Native Americans. Since this is a large task for a brief space, I will focus on main points. I begin with general remarks and end by applying them as brief criticism of the papers just presented.

One should expect that the forms of discrimination would vary across minority groups during a given year and during a given century. Some differences between blacks and Native Americans are so important that policies or predictions based on intuition developed with current models of discrimination against blacks lead to serious error when applied to Native Americans. One error is to omit land and sovereignty as variables. Another is to fail to examine the full range of policies open to the majority when they discriminate.

Because Native Americans have always owned large amounts of land and still do, an exclusive focus on labor markets is incorrect. The nature of property law and the limitations on it are different for land, physical capital, and human capital. As a result, the benefits and costs of different forms of discrimination change as compared to blacks. Hence, the preferred forms of discrimination also change.

For Native Americans, one must emphasize the benefits of sovereignty and focus upon the problems which tribal governments have protecting and utilizing their constitutional status. Personal sovereignty, or freedom, is emphasized for blacks. Governments should be able to tax and to control property law among individuals within their jurisdiction, including whites who settle upon confiscated lands. Native Americans ruled themselves originally and have attempted to adjust their political institutions in order to retain control over internal affairs, in spite of becoming politically subordinate to the United States government. I believe that the power to tax and to write civil and criminal law are important economic variables.

Although land and sovereignty are key variables, the sad fact is that Indians have suffered continual erosion in their possession of both. With more space, I would elaborate the nature of the erosion. Instead, I shall indicate how these two variables change models of discrimination.

Suppose a tribe owns land whose products such as timber, fur, buffalo, minerals, or oil can be transformed into capital with little net borrowing. Suppose, further, that historically the tribe has successfully administered property rights in land among its members, with exclusive use-rights but a system of inheritance different from the American norm. Suppose, also, that it can tax all, white or Indian, who live within a designated boundary. Then questions about segregation in work places or residences are not major. More important are questions about trade with outsiders, immigration and intermarriage, and the ability of the tribe to prevent confiscation of land and sovereignty.

The federal government has always had a conflict of interest in its chosen role as protector of Indian property rights. American politicians and bureaucrats perceive a choice in the implementation of discrimination between trade methods and confiscation methods. Let us examine briefly the benefits and costs which affect the choice. Thinking in terms of centuries, the benefits of confiscation vary with the price of land, with land classified by type of natural resource. The costs depend upon other factors. For instance, in 1974 the strength of the sympathy vote is important. In 1874, the availability of rifles, ammunition, and horses mattered. In 1774, the strategic relationships between a tribe and the British mattered. At times, the American government has preferred trade to confiscation after examining the benefits and the costs.

Finally, I must emphasize that Native Americans are very diverse. Our cultures are different from one another, both aboriginally and in the type of borrowing from Europeans. Further, the relative success in retaining land and sovereignty varies greatly among tribes. Those who wish nationally valid generalizations risk error should they try to apply the same analysis to the Native Americans of Alaska, New York, Montana, and Washington; these four states, for instance, each represent four different legal situations.

Currently, the short-term outlook for Native Americans is good. A year ago Congress restored the sovereignty of the Menominee Indians of Wisconsin by reversing the abolition of their reservation. Congress rescinded a law passed in 1954 and implemented in 1961. In 1971 Congress returned some land to the Taos Pueblo in New Mexico. The Yakima and Warm Springs Reservations have also had land returned. How long will this policy last? Native Americans do not now assume that their position is secure. The federal government currently finds a trade policy preferable to a confiscation policy. Perhaps the benefits and the costs which led to this decision will change again in the future.

I disagree with the previous speakers in their choice of variables to exclude when discussing future prospects. When David Swinton spoke of resources, he usually meant human capital, not land or even physical capital. He did not discuss sovereignty. If he is right to conclude that the market currently has no inherent tendency either to lessen or to increase inequality, I would suggest examining the potentials of confiscation. If he had concluded that the market currently tends to lessen inequality, I would have asked if confiscation was imminent.

Robert Fogel and other scholars examining slavery recognize the choice between money and force to accomplish the transfer of resources to whites from nonwhites. As in the North, whites in the South displaced Indians from their land before 1860. Those Native Americans joined others in Oklahoma. There they retained substantial sovereignty until early in this century when they endured another onslaught of confiscation and were finally reduced to poverty.

Bernard Anderson and Phyllis Wallace also focus exclusively on labor markets. I would emphasize that Native Americans experience high levels of unemployment and limited access to high-paying occupations in addition to the problems I referred to earlier. Policy based on exclusive attention to these employment problems, however, leads to serious errors. It is for this reason that I stressed land and sovereignty in these remarks.

Many of you may feel that these variables are also important for other minorities. Although I agree, the diversity among Native Americans makes me expect much variation in the manner in which land and sovereignty affect other minorities.

DANIEL FUSFELD, University of Michigan: I was asked to comment on the last two papers. I did not obtain Robert Fogel's paper until this session, so I cannot really comment on it.

Regarding the last two papers, by Bernard Anderson and Phyllis Wallace, and David Swinton, I did not like either one, in many ways, for much the same reasons. They both stay within the establishment, neoclassical economic framework. That is just the wrong way to approach the problems of minority groups, particularly those of blacks, Indians, and women. The neoclassical model assumes continuity in labor markets and unified labor markets; it assumes even adjustments across the board, and it excludes exactly the kind of relevant variables that Fogel, for example, included in his analysis. The key point that Fogel makes is that the viability of the large-scale plantation in the South before the Civil War was due to a high degree of institutionalized exploitation of black labor. It was not because large-scale plantation agriculture was more efficient than the small farm. If it were more efficient, we would have had a recurrence of large-scale plantation agriculture after the Civil War, which we did not get. That mode of economically successful operation was made possible by exploitation, by the use of force and power to exploit blacks.

Another point about Fogel's analysis illustrates why the neoclassical analysis cannot help us understand exploitive patterns of behavior. Those plantations were efficient as economic units. People running them sought to maximize their welfare, as did the slaves, given the opportunities available. Yet it was far from a social optimum. The central problem involves the basic institutional parameters within which the profit-maximizing activities of individuals or other economic units are carried on.

Let me develop this point specifically with respect to the other two papers. Anderson and Wallace present a view of the economy as a whole. The welfare of blacks and other ethnic groups is said to have improved during the 1960's, and to have deteriorated maybe in the few years since then. The problem here is—which blacks are we talking about? Do we consider the black man who picks up the trash in front of my house every week, or the black psychiatrist who lives next door? We may very well hire black, professional economists to teach students at the University of Michigan or other institutions of higher learning. This may have no impact whatsoever on the way of life of that poor guy who picks up the trash. Yet, hiring the black economist to teach in the university may statistically pull up the average. Thus, we must not look at just the overall view of the labor market, confining ourselves to averages. We need to go beyond what Anderson and Wallace did and look at it in terms of the cyclical aspects, prosperity versus not so prosperous times, and rapid growth versus stagnation. We also have to consider what is transpiring within the minority community. There is good reason to believe, on both empirical and theoretical grounds, that, while the averages showed an improvement in the condition of blacks taken as a whole during the 1960's, there was still a deterioration in the economic and social condition of blacks living within the inner city ghettos. It is that kind of a question, it seems to me, that we have to get into in much greater detail.

Turning to Swinton's paper, the line of argument was basically this: first, the economic position of minority groups relative to the majority seems to have stabilized within a new economic equilibrium that has emerged after the urban revolts of the 1960's. Second, there is little reason to believe that economic forces alone will eliminate the present disparities that seem to have stabilized. And finally, the only alternative is political action that will widen opportunities, redistribute resources, and reduce discrimination.

I agree fully with that line of argument, but I do not think Swinton properly analyzed why economic forces alone will not be able to eliminate the present disparities. He was basically saying that disparities between the races are due to disparities in the level and structure of opportunity, and those disparities are created by discrimination against minorities. Likewise, there are disparities in the distribution of resources which lead to further economic disparities, but those are themselves related to discrimination. The entire argument is circular: the economic system alone will not correct the problem of discrimination because of discrimination.

Another part of his argument is that there are inflexibilities and imperfections in the way in which the market functions; many of the disparities between minorities and majorities are the results of these imperfections. If I might step out of my normal abode and become an apologist for the economic system, the answer is obvious: all we need to do is make a few marginal adjustments in the economic system to get rid of the imperfections; and we then need to change the majority by educating it a little bit to get rid of its propensity to discriminate; and maybe at the same time provide some more human capital to minority groups; and then let the economic system slowly and gradually but completely eliminate the disparities. His argument does not show that the economic system, as such, will fail to terminate the inequality. It is the things wrong with the economic system that are at fault. A little action here and there can set things right. The nonclassical argument remains.

I would like to cast the discussion in a completely different framework. Returning to Fogel, we can state that coerced labor has always been a feature of the American economy, particularly with regard to a variety of minority groups: blacks, Indians, and women, in particular. With respect to blacks, before the Civil War there was the system of slavery, followed after the Civil War by a system of debt-tenure in southern agriculture in the sharecropping system, in turn followed after the breakup of the sharecropping system with ghettoization and crowding of blacks into menial occupations. The economy moved from a legally sanctioned form of coercion to a quasi-legal form, to what is now essentially an economic and social form of coercion, far more subtle in its operation, all three having the key economic characteristic of excluding the minority groups from some opportunities which then enables them to be exploited in the occupations or employments into which they are crowded by slavery, sharecropping, or ghettoization. If this is the economic history of blacks and other minority groups—in this case, particularly blacks—note one very fundamental point: each transition was marked by a major social, political, or economic trauma following the period of stability and viability of the particular forms of coercion of labor that existed. Slavery as a form of coerced labor was ended with the Civil War, certainly a traumatic event. Sharecropping and its system of debt-tenure was ended by a series of economic disasters starting with the boll weevil, including the impact of the Great Depression, and ending with the mechanization of the labor process in southern agriculture. This series of drastic events coming from essentially outside the labor system itself was much more traumatic for blacks than for any other group.

The question is then raised: given the proposition that the present system of ghettoization and coerced labor has been stabilized in the institutional structure, what kind of traumatic economic or political event or set of events is going to change it? I really agree with Swinton. He is right. The present system has been very largely stabilized. It is not going to be broken by economic evolution. The kinds of things that are going to break it, however, it seems to me, are going to be far more drastic than, as he suggests, political action by the minority groups themselves.

THE HUMAN CAPITAL APPROACH: AN APPRAISAL

Human Capital Theory: Education, Discrimination, and Life Cycles

By FINIS WELCH*

The most important point I can make in assessing contributions of human capital theory is that it is a tool for analysis and is neither a religion nor a shrine. Like any other analytical tool, we use it when it simplifies interpretation of complex phenomena and reject it when it conceals more than it reveals. Much of the burden of defining "what human capital theory says" has been assumed by detractors who, in many cases, offer caricatures in lieu of characterization. Omissions and abstractions are criticized without regard for the context in which assumptions are made and without reference to the predictive power of simplifying assumptions. Human capital theory is based only on the presumption that labor skills or market perceptions of skill are both durable and malleable. This, surely, is not a philosophy of labor markets.

Three topics are discussed here. They do not represent a comprehensive view, but are each part of the overall question of income distributions and have each absorbed the attention of many working with human capital theory. The topics are: (1) wealth distributions and life-cycle earnings profiles; (2) sources of income returns to schooling; and (3) race differences in income. My objective is to describe what I think we have learned about these phenomena as well as to point to the most important unsettled issues.

I. Wealth Distributions and Life-cycle Earnings Profiles

The popular theories of G. S. Becker (1965), Y. Ben-Porath, J. Mincer (1962 and 1974), and S. Rosen view the distribution of labor earnings over a worker's career as endogenous to his own behavior. Although objective functions are usually described in terms of full life-cycle utility, in application the objective is maximization of discounted earnings. Work serves a dual function: it contributes current earnings and affects future productivity. Workers are viewed as determining the rate at which current earnings are sacrificed in light of future prospects. Since the period over which earnings can be recouped declines as age increases, the value of currently foregone earnings declines with increasing age. This prediction of declining investments with increasing age is the most consistent feature of the standard theories, although they also allow the rate of investment in future productivity to depend on the extent to which previously acquired skills enhance or detract from the ability to continue the skill acquisition process. Because investments decline with age, these theories invariably predict observed earnings profiles that possess negative second derivatives over much of the worker's life.

We can lament the fact that none of these theories have been reduced to simple, i.e., empirically tractable, and identifiable functions describing lifetime earnings. Such functions would be helpful in evaluating impacts of policy options ranging from income maintenance programs to subsidized training and unemployment insurance. Yet the lack of available data for lifetime earnings profiles of individuals not only limits us in identifying such functions, but more importantly, it prevents evaluation of the quantitative impact of investments in human capital in shaping earnings. Without time series observations of individuals, we have turned to analyses of cross-sectional data which compare different people of different ages.

---

* University of California at Los Angeles and the Rand Corporation. I am indebted to Dennis DeTray and James P. Smith for comments during the construction of this draft.

Thanks largely to Mincer (1974) we have evidence that in the cross-section, earnings profiles appear very much *as though* they result from investment processes like those described above. This is true not only of mean earnings at various ages for persons of similar schooling, race, etc., but earnings variance among individuals appears to exhibit U-shaped age profiles. The variance idea is central to contrasts of wealth, i.e., discounted lifetime earnings, distributions with earnings distributions at various points in the life-cycle. Suppose that within a given schooling-race cohort individuals faced with differing ability, tastes, access to investable funds, etc. elect differing time shapes for their own earnings profiles. Early in the career, heavy investors will have low earnings (relative to their alternatives) which are subsequently compensated by high earnings. Given a distribution of investment profiles among individuals, it seems reasonable to expect age-related earnings profiles to be dominated by differences in investments early in the career. But in the early phases, earnings variances will decline with age as investments and returns on earlier investments tend toward zero sum. Still later, variance will rise as returns to earlier investments dominate. It is this view which leads Mincer to predict U-shaped age-earnings variance profiles, and these predictions are confirmed in his analysis of the 1960 census.

At any point in the worker's career, earnings can be viewed as the sum of two terms. One corresponds to his permanent income—the average annual equivalent value of his discounted lifetime earnings. The other, his net investment, is the sum of currently foregone earnings as investments in the future and the realized return on his past investments. This second term has zero discounted value over the life cycle, and although it may add or detract from earnings variance among individuals at any point, it plays no role in comparisons of lifetime wealth. Unfortunately, available data permit only guesses about the numerical import of that part of earnings variance at a point in the career which is fully compensated over the life cycle. Lee Lillard has analyzed data from the National Bureau of Economic Research (*NBER*)-Thorndike sample. It contains observations of approximately 5,000 individuals with earnings at five career points. Lillard first estimates a relatively unconstrained function and finds, as did Mincer, that the shape of the estimated earnings function is consistent with the investment view. He next contrasts a measure of average annual earnings variance with alternative estimates of variance in the distribution of discounted lifetime earnings under two extreme assumptions about the life-cycle persistence of residuals observed for individuals. The lower estimate assumes that residuals are purely transitory and under this assumption the coefficient of variation for discounted lifetime earnings is only about one-fourth of the coefficient of variation for mean earnings.[1] The upper bound assumes that residuals are completely persistent throughout the worker's life. Under this assumption, the coefficient of variation for discounted earnings is 40 percent greater than the corresponding measure for mean earnings. Lillard felt that, with only five data points for each individual, he was unable to make more refined statements about the autocovariance structure of individual residuals from the estimated earnings structure. Nonetheless, his estimates serve as an indication of our ignorance, for the range between 0.25 and 1.4 is large and we have little information about the role of compensating variations in observed earnings distributions.

---

[1] Lillard's lower estimate is not a lower bound since it is possible, but unlikely, that residuals from his estimated equation are not transitory at all but are fully compensated within the life cycle. In this case, the coefficient of variation for wealth is roughly 10 percent of the coefficient of variation for annual earnings.

Mincer's prediction of a U-shaped age-earnings variance profile hinges on the assumption of an "overtaking age" that has reasonably small variance among individuals in a cohort. The overtaking age is defined as the age at which observed earnings are equal to average annual equivalent earnings for the discounted lifetime profile. Based on the notion of a common overtaking age for members of a cohort, John Hause reports data both for Sweden and the United States which show significant negative auto-covariance among individuals. This is evidence of compensating variation, but unfortunately it has not been quantified in a manner useful for reducing the range on Lillard's bounds. It can only be taken as a hint that the upper bound may be a significant overstatement.

To me, the question of the importance of the human capital concept in interpreting differences in human wealth is closely linked to the question of the numerical import of compensating variations. Yet resolution of this question awaits either better data or more ingenious use of existing files.

II. Sources of Income Returns to Schooling

The positive association between income and schooling is well-known, but the nature of the underlying mechanism continues to be a topic of lively debate. The early literature of human capital assumed, perhaps blithely, not only that a causal relation exists, but that the zero-order association that corrects neither for "ability" nor "background" offers an adequate basis for calculations of rates of return and for estimates, based on these returns, of the contributions of increased schooling levels to economic growth. Recently a number of authors have questioned assumptions on which the early calculations were based. The dissent is wide-ranging and includes questions of the role of ability and family background and the nature of acquired skills. It also includes the fundamental question of whether schools contribute directly to market skills or only serve an identifying or screening function. My purpose is to summarize some salient features of this discussion along with the available evidence.

It seems noteworthy that the fundamental notion of human capital, of foregoing current income for the prospect of increased future earnings, assumes only that the schooling-income association is not spurious. As such, it is fully consistent with the screening view that schools primarily identify preexistent skills and with the view that market skills are produced in school. If marketable skills are produced in school, it does not matter from a positive perspective whether the skills are predominately affective as H. Gintis posits or whether they are cognitive.

Is the income-schooling association spurious? Here, on the basis of available data, the answer is no. There are numerous available studies that consider earnings as functions of schooling, measured ability, and family background. These studies consistently show that the partial relation between income and schooling is weaker, i.e., implies lower returns to schooling than the zero-order association. While there is a wide range of reductions in measured returns to schooling that follow corrections for ability and family background, none of the estimates show a zero net return to schooling.

Briefly, consider the evidence. Gintis summarizes nine different studies in which the reduction in the schooling coefficient achieved by an ability correction ranges from 4 to 35 percent and in these studies the central tendency appears to be a little above 10 percent. E. Hanushek partitioned a large data base into 200 subsamples based on race and region of residence. The average schooling coefficient is reduced by about 15 percent by introducing an ability correction. P. Taubman and T. Wales report a 9 percent reduction in the schooling coefficient when correcting for a measure of general ability and a reduction of 30 to 35 percent when correcting for mathematical ability.

The paper by Z. Griliches and M. Mason is particularly interesting in these comparisons. Their task is not only one of assessing the sheer size of the school and ability effects, but of analyzing robustness under specifications that alternate control for factors pertaining to family background, race, and residence and then leave these controls free to vary so that their effects are captured by correlates. A unique feature of the Griliches-Mason data is that ability, the armed forces qualifying exam, is observed at an intermediate point in the schooling career—the point of entry into the armed forces. In principle, this allows identification not only of effects of measured ability on income, but of biases in returns to schooling due to a failure to correct for initial ability. Their estimates show a reduction in schooling's measured contributions to earnings of 7 to 10 percent when controlling for measured initial ability and of reductions of 13 to 17 percent when controlling for postschool measured ability. Gintis does not distinguish in his survey whether ability is measured pre- or postschool. Hanushek's measure is postschool and Taubman and Wales use a measure obtained on completion of high school for a sample in which the majority of observations are college graduates.

Just as measured ability and school completion are correlated, so are family socioeconomic status (*SES*) and schooling. As with ability corrections, most studies controlling for family background report small reductions in estimates of income returns to schooling. J. Morgan and M. David note that including variables for father's education and an index of family *SES* reduces the measured contribution of schooling to income by 12 percent. F. T. Juster reports a slightly smaller effect. And M. Carnoy (1964) using data from Mexico finds independence between computed rates of return and father's occupation. In a sample restricted to persons of very high measured I.Q. (above 135), Arleen Leibowitz finds no relation between family *SES* and income, but does find that school completion and family *SES* are positively correlated. S. Bowles, on the other hand, estimates that controlling for family background reduces the apparent return to schooling by 40 percent.

In this context, the Griliches-Mason work is also illuminating for it is not clear how the interdependence between schooling, income, and family background should be interpreted. Family background may serve as a proxy for quality of schooling and for complementarities between school and home environments or it may index access to good, i.e., high salaried, jobs for which schooling is necessary but not sufficient. If background indexes the quality of the learning environment, we would expect effects of family background on income contribution to schooling to be higher for schooling acquired while the child resides with his parents than for schooling acquired after leaving home. In the Griliches-Mason data, the school career is interrupted by service in the armed forces, and they find that by controlling for family background the income contribution of schooling is reduced by 25 percent for pre–armed forces schooling and by only 3 to 6 percent for schooling acquired after being in the armed forces.

Controls for ability and/or for family background appear to reduce the measured contribution of income to schooling but they do not reduce it to zero. The argument that the income-schooling relationships are causal remains intact.

The available literature either ignores or treats only lightly the question of measurement error. Yet I think this is a problem of considerable importance. Frankly, I find it hard to conceive of a poorer measure of the marketable skills a person acquires in school than the number of years he has been able to endure a classroom environment. My only justification for using such a crude measure is that I can find nothing better. Too, I am very agnostic about what "ability" tests measure.[2] Similar statements hold for measures of home environment. Yet as soon as we open the door to measurement error, we find a completely nebulous world in which almost anything is possible.

Consider a simple regression model of measurement error described as:

$$y = \beta_1\tilde{s} + \beta_2\tilde{a} + u \tag{1}$$

$$s = \tilde{s} + v_1 \tag{2}$$

$$a = \tilde{a} + v_2 \tag{3}$$

with the residuals or measurement errors, u, v_1, and v_2 being mutually independent as well as independent of $\tilde{s}$ and $\tilde{a}$. Here y denotes individual income, $\tilde{s}$ is marketable skills acquired in school scaled to the units of school years completed, and $\tilde{a}$ describes true initial (preschool) ability. Observations of schooling and ability, s and a, are with error, v_1 and v_2, respectively. It is easy to show that:

$$\textit{plim}\, b_{ys} = (\beta_1 + \beta_2\rho_{\tilde{a}\tilde{s}})r^2_{s\tilde{s}} \quad \text{and} \tag{4}$$

$$\textit{plim}\, b_{ys\cdot a} = \beta_1 r^2_{s\tilde{s}\cdot a} + \beta_2\rho_{\tilde{a}\tilde{s}}r^2_{s\tilde{s}}(1 - r^2_{a\tilde{a}\cdot s}) \tag{5}$$

where *plim* refers to the limit in probability, b_{ys} is the ordinary least square (*ols*) coefficient from the regression of y on s, $b_{ys\cdot a}$ is the *ols* coefficient of the regression of y on s holding a constant, $\rho_{\tilde{a}\tilde{s}}$ is the *ols* coefficient from the auxiliary regression of $\tilde{a}$ on $\tilde{s}$, $r^2_{s\tilde{s}}$ is a squared zero-order correlation coefficient and $r^2_{s\tilde{s}\cdot a}$ and $r^2_{a\tilde{a}\cdot s}$ are squared partial correlations.[3] It is assumed that $\beta_1, \beta_2, \rho_{\tilde{a}\tilde{s}} > 0$.

Under these assumptions, it is not clear, a priori, that the regression which does not correct for ability overstates the income return to schooling. It is understated (in the probability limit) by $\beta_1 (1-r^2_{s\tilde{s}})$ and overstated by $\beta_2\rho_{\tilde{a}\tilde{s}}r^2_{s\tilde{s}}$. Similarly, it is not clear that "corrections" for ability improve the estimate of β_1, the contribution of schooling to earnings. Comparison of the uncorrected, b_{ys}, and corrected, $b_{ys\cdot a}$, measures show unambiguously that (in the limit) $b_{ys\cdot a} < b_{ys}$. The correction has two effects. It reduces the overstatement from $\beta_2\rho_{\tilde{a}\tilde{s}}r^2_{s\tilde{s}}$ to $\beta_2\rho_{\tilde{a}\tilde{s}}r^2_{s\tilde{s}} (1-r^2_{a\tilde{a}\cdot s})$ but since $r^2_{s\tilde{s}\cdot a} < r^2_{s\tilde{s}}$ *it increases* the understatement from $\beta_1(1-r^2_{s\tilde{s}})$ to $\beta_1(1-r^2_{s\tilde{s}\cdot a})$.[4]

My point is that part of the recent literature seems to have evolved into a game in which we search for the set of control variables which minimize the measured direct return to schooling. I hope it is obvious from the errors-in-variables discussion that this game is not without pitfalls.

---

[2] I have summarized these objections elsewhere and will not repeat them here (see Welch 1974b).

[3] All correlations here refer to limits in probability.

[4] In this description: $r^2_{s\tilde{s}} - r^2_{s\tilde{s}\cdot a} = r^2_{sa}(1 - r^2_{s\tilde{s}\cdot a}) > 0$.

What can be said about the nature of skills associated with schooling? In this, as in most areas of economics, the data offer only a few hints. Gintis, noting that measured returns to schooling corrected for postschool ability average 80 to 90 percent of the uncorrected measures, argues that market skills associated with schooling are not the cognitive skills that tests measure. This interpretation seems valid, but due to the errors-in-variables discussion, I would underscore that the income return appears not to consist mainly of what ability tests *measure*. Elsewhere I have argued that, in dynamic environments in which abilities to acquire and decode information are at a premium, the return to schooling may partly reflect allocative abilities.

Carnoy's finding (1967) that estimated returns to schooling in Latin America are positively associated with rates of growth in real per capita income is relevant here. I have offered similar evidence showing in *U.S.* agriculture that wages of persons who have attended college are higher relative to others in states where the rate of productivity advance is most rapid. There are by now a number of available papers that report evidence of links between allocative efficiency and schooling in agriculture.[5] These studies include data for the United States, Brazil, Colombia, India, and Taiwan, and to me the cumulative impact is convincing. Unfortunately, except for the Carnoy paper, we do not have evidence of whether similar relations hold outside agriculture. For other industries, we remain largely ignorant of the role of education in production.

What of schooling as a screening device? The basic argument is that part of the role of schools is the identification of preexistent skills. If an individual's productivity does not depend upon the market's perception of his ability and his wage does, this possibility drives a wedge between private and social returns to schooling. The idea that an individual's product is independent of assessments of ability is puzzling. It seems reasonable that a person with high ability who is perceived as having low ability would be assigned tasks in which he would be less productive than he might otherwise be. Nonetheless, if both workers and employers are ignorant of individual ability and if schools serve an identification role, I find it hard to conceive of market experiments which distinguish this from a world in which schools produce rather than identify skills. Markets presumably reward skills independently of their origin.

The model spelled out by M. Spence is different. In it, workers initially know their abilities. Employers are ignorant and turn to schools for estimates of ability. K. Wolpin attempts to test this model by contrasting schooling levels and estimates of income returns to schooling between self-employed and private wage and salary workers. He finds, for the *NBER*-Thorndike sample, similar school completion distributions and similar estimates for the return to schooling for these groups. The novelty of this approach is that, in a world where the income return to schooling is derived exclusively from schooling's role in supplying information to employers, we are immediately suspicious of incentives for those who will be self-employed to go to school. It is not surprising that measured returns to school completion are equal for the two groups for more schooled workers are posited by the screening models as having more ability. Yet, one wonders why self-employed workers go to school at all if their productivity is not enhanced by the process.[6]

---

[5] The authors include: N. Khaldi, G. Fane, C. Wu, S. Sidhu, T. Haller, F. Patrick and E. Kehrberg, Welch (1970) and W. Huffman.

If I can make an overall criticism of research into income-schooling relationships, it is that it seems superficial. For too long we have proceeded as though there is something intrinsically meaningful about the number of years of school a person has completed. Except for the model of Ben-Porath and the solutions presented by W. Haley and Lillard, we do not have clear statements of the endogenicity of schooling vis-à-vis ability. Most importantly, we have not been able to generalize in regard to the role of family background as it contributes to motivation, to the "quality" of the learning environment, and to access to funds for investments in human capital. Similarly, I cannot point either to a single cogent discussion of the various dimensions of ability or to empirical evidence pertaining to determinants of differential values for different kinds of ability. We remain very ignorant about the impact of quality of schooling and of its various dimensions. We presume that somehow it matters what schools do. But what they are doing varies within schools and between schools at a point in time and through time.

It seems that a full description of individual differences in earnings capacity would answer these questions: (1) Within a market, what is the nature of the various human skills that affect earnings? (2) What determines variations in the level of remuneration for given skills between markets? (3) What factors determine the distributions of skills among individuals? Given our ignorance of these questions, I do not think we can be very sanguine about the accomplishments of labor economics in general and of human capital economics in particular. Human capital approaches have highlighted schooling and all that it conceals as "the" skill of interest. At the expense of this myopia in regard to the first question on the above list we remain largely ignorant of answers to the second and third.

III. Race Differences in Earnings

In the last two decades there has been a proliferation of economic theories of discrimination and of empirical studies which measure racial differences in earnings. Yet in this, more than any area I know of, the dialogue between theory and data has been minimized.[7] If there is blame, it seems to me that the major fault should be assigned to the theories, for they have failed in describing what the data show. I take four empirical "facts" as given. One fact is that, throughout the period for which data are available, black earnings in the United States have been less than white earnings. The second is that black-white differences in earnings have, at least in the recent past, been declining. The third is that black-white skill differentials are declining. And the fourth is that labor markets are increasingly integrated.[8] To this list, I would add that, historically, income returns to schooling have been less for blacks than for whites but recently, at least for those with "some college," returns to schooling are at least as high for blacks as for whites (Welch, 1973, or Smith and Welch).

---

[6] There are, of course, taste and insurance factors which would enter the schooling decision. But whatever the dimensions of the decision, schooling should be less attractive to those who get no directly marketable skills from it.

[7] I exempt the papers by W. Landes on fair employment legislation and R. Freeman on "public" discrimination from this criticism.

[8] Evidence of the first point seems unnecessary. However, there is some recent data (A. Horowitz) on females showing that during the 1960's earnings of black females surpassed earnings of white females. These data are highly aggregative and it is unclear whether the result will withstand decomposition by age, schooling, and work experience. For the second point, see the analysis of the *U.S.* Censuses in 1960 and 1970 by J. Smith and Welch, and the papers by Horowitz and W. Vroman. The third point can easily be verified if confidence is placed in schooling as a nominal skill index by any aggregated data source. Finally, the fourth point is largely a guess but is backed by evidence of convergent occupational and industrial employment patterns shown in the 1960's via census comparisons and in O. Ashenfelter's analysis of unions in the United States.

Now consider theories of discrimination. The original work of Becker and extensions by K. J. Arrow focus mainly on taste.[9] Both workers and employers are presumed to care about the race composition of the work force within firms. Consumers, on the other hand, are presumed indifferent about the race composition of labor employed in production of impersonal products. From this it follows immediately that employers who discriminate by paying higher wages to whites, whom they prefer, than is necessary in view of opportunities for hiring similarly qualified blacks do so at the expense of nominal profits. The long-term instability of such models is obvious. I know of no empirical attempt to verify that profits of firms depend upon the racial composition of their work force. Further, I think this omission itself is evidence that we fail to take seriously the most obvious implication of employer discrimination.

Discrimination by employees is more interesting. If employees require higher wages to work in integrated firms, then discriminatory equilibria are possible. They are not possible if all laborers are perfect technical substitutes, for in this case the theory degenerates into a prediction of completely segregated production. But if blacks and whites are not perfect substitutes, integrated equilibria can emerge in which the marginal cost of hiring a worker includes his wage and may also include the cost which the firm realizes in having to compensate other workers for whatever psychic losses they incur as a result of adding the worker under consideration. But what are the longer term implications of such equilibria?

Suppose that there are only two types of labor, skilled and unskilled, and suppose that production is a linear homogeneous function of these inputs. If the distribution of whites between skilled and unskilled labor is the same as for blacks, the resulting market equilibrium will be one of complete segregation without wage discrimination. If, on the other hand, a higher proportion of black labor is unskilled, partially integrated production may result. This is not a necessary consequence for, as Stiglitz notes, whites may specialize disproportionately in production of skill-intensive products and trade between products can substitute perfectly for integration.

But suppose there is only one product. In this case, neither skilled blacks nor unskilled whites would work in integrated firms. Some unskilled blacks and some skilled whites would be integrated to the extent that the complementarities of joint production compensate for diseconomics of association. The resultant equilibrium would entail higher wages for the relatively scarce, skilled black workers than for skilled white labor and lower wages for unskilled blacks than for unskilled whites. Over time the ratio of skilled to unskilled black labor rises relative to that of whites in response to the premium on black skill and, consequently, the extent of integrated production declines.

I suspect that the static predictions of this "theory" are historically satisfied in the United States. On balance, I suppose that whites have been disproportionately employed in skill-intensive industries. I also suppose that there has historically been more integrated production involving skilled whites and unskilled blacks than has involved integration of skilled blacks with unskilled whites. But it seems to me that the dynamic predictions vis-à-vis the extent of integration have been contradicted. Certainly the prediction of convergent skill distributions is satisfied, but has skill convergence resulted in increased segregation? I do not think so. Historically the bulk of black college graduates has entered professions that largely service black clientele. More recently black college graduates are entering managerial and engineering professions, and I am convinced that this change has been coincident with increased integration.

---

[9] In this section, I draw heavily from the excellent summary provided by J. E. Stiglitz.

Economic theories of discrimination, at least the taste-based theories, have tended to be theories of segregation rather than theories of wage differences, and their predictions of rising segregation through time are not supported empirically. On the other hand, empirical analyses of discrimination focus almost exclusively on wages. Race differences in wages are partitioned into direct and indirect components. Indirect effects refer to differentials associated with worker characteristics, such as schooling, and direct effects refer to race differences in parameters of estimated regression equations. This kind of partitioning is valuable since it offers information relevant to understanding origins of discrimination. But clearly the estimated direct effects do not fully "measure" discrimination: questions of underlying causes of differences in characteristics are also important.

For example, human-capital-styled theories focus both on the costs and returns to investments in skill. Discriminatory theories predict higher returns to skill acquisition for groups in which skill is most scarce, but empirical verification of the existence of these premia founder on the question of the quality of schooling. Although we remain largely ignorant about ingredients of school quality, speculations about the impact of race differences in quality of schooling abound. The reasons are obvious.

First, discrimination in publicly provided services is easier to understand since costs to discriminators are less apparent than is true of market discrimination. Second, insofar as school quality is concerned, there are implications both for costs and returns. If we ignore market discrimination and assume that private costs of schooling are described fully by the opportunity costs of a student's time, it is clear that returns as compared to the costs of schooling are higher where quality is higher, i.e., where the public subsidy is greatest. This comparison is flawed both by market discrimination and by the possibility of productivity differences that would exist independently of schools. But it is clear in any case that school quality is a legitimate concern, and the historical record of large race differences in attributes of schools is evident.

I, for one, do not think we have gained much insight of discriminatory mechanisms either from theories of discrimination or from measures of wage differentials. In the early discussion, public discrimination was dismissed. It was too obvious and did not tax our intellectual capital. I think we should reexamine this earlier omission. The historical experience with black schools in the United States offers a point of departure. Following disfranchisement in the last quarter of the nineteenth century, quality of black schools deteriorated reaching a nadir sometime in the early part of the twentieth century.[10] Then black schools began to compare increasingly favorably to white schools. This improvement was clearly under way in the 1930's and continued at least through 1960, without an apparently significant deviation from trend in 1954, the time of the desegration decision. If we could explain reasons for the decline following disfranchisement and the subsequent steady improvement that, significantly, led refranchisement, I think we would have at least begun to understand public discrimination.

---

[10] Here, by quality I mean any of measurable attributes of schools: expenditures per pupil, teacher salaries, retardation rates, days attended per year, teacher/pupil ratios, etc. See Welch (1974a).

Regarding market discrimination, there are a number of important unanswered questions. I am, however, not sanguine about prospects of data adequate for their resolution. One involves the vintage life-cycle nexus. Distinctions can only be made with observations through time, but then we must be able to reconcile these data with effects of market fluctuations. Another involves questions of effects of "statistical" discrimination. How closely is one tied to the mean performance of people "like" him? Of particular concern is the issue of self-fulfilling prophecies. Until we gain information about ties between true productivity and perceptions of productivity, we will remain ignorant about what may be a fundamentally important source of discrimination. Suppose an individual's actual productivity is bounded by perceptions of his ability: workers assigned to highly routinized tasks cannot demonstrate originality or imagination. In such a world returns to individuality to a person distinguishing himself from his peers may be slight in comparison to costs. And prophecies of group productivity differentials may become self-fulfilling.

IV. Concluding Comment

I have discussed only three of an array of topics that concern those who work with human capital theory. Life-cycle earnings patterns are the core of this theory while underlying sources of income returns to schooling and race differences in wages are not necessarily related to it. In fact, I know of no "human capital theory of discrimination." Theories of human capital and of discrimination tend to be associated because both are parts of a comprehensive view of earnings differentials and because work by Becker serves as a benchmark in each area. The most recent developments in human capital theory include the new home economics and work on the economics of human fertility. It is somewhat early for evaluation of this effort, but the subjects are important and the potential is high. In any case, their point of departure consists of recognizing the fundamental role played by the cost of human time in the allocation of time. And this, more than any other insight, may be the most valued contribution of human capital theory.

REFERENCES

K. J. Arrow, "Models of Job Discrimination," in A. Pascal, ed., *Racial Discrimination in Economic Life*, Lexington 1972, *1*, 83–102.

G. S. Becker, *Human Capital*, Nat. Bur. of Econ. Res., New York 1965.

———, *The Economics of Discrimination*, Chicago 1957.

Y. Ben-Porath, "The Production of Human Capital and the Lifecycle of Earnings," *J. of Polit. Econ.*, Aug. 1965, *75*, 352–65.

S. Bowles, "Schooling and Inequality From Generation to Generation," *J. of Polit. Econ.*, Mar./Apr. 1972, *80*, Supp., 219–51.

M. Carnoy, "Rates of Return to Schooling in Latin America," *J. of Human Resources*, Summer 1967, *2*, 359–74.

———, "The Cost and Returns to Schooling in Mexico: A Case Study," unpublished doctoral dissertation, Univ. of Chicago 1964.

G. Fane, "The Productive Value of Education in Agriculture in the U.S. Corn Belt, 1964," unpublished doctoral dissertation, Harvard Univ. 1972.

R. Freeman, "Labor Market Discrimination: Analysis Findings and Problems," in M. D. Intriligator and D. A. Kendrick, eds., *Frontiers of Quantitative Economics*, Amsterdam 1974, *2*, 501–55.

H. Gintis, "Education Technology and the

Characteristics of Worker Productivity," *Amer. Econ. Rev. Proc.*, May 1971, *61*, 266–79.

Z. Griliches and M. Mason, "Education, Income and Ability," *J. of Polit. Econ.*, Mar./Apr. 1972, *80*, 74–103.

W. Haley, "Human Capital: The Choice Between Investment and Income," *Amer. Econ. Rev.*, Dec. 1973, *63*, 929–44.

T. Haller, "Education and Rural Development in Colombia," unpublished doctoral dissertation, Purdue Univ., June 1972.

E. Hanushek, "Regional Differences in the Structure of Earnings," *Rev. of Econ. and Statist.*, May 1973, *55*, 204–13.

J. C. Hause, "The Covariance Structure of Earnings and the On-The-Job Training Hypotheses," unpublished paper, Univ. of Minnesota, August 1974.

A. Horowitz, "Pattern and Causes of Changes in White-Nonwhite Income Differences: 1947–1972," in G. M. Von Furstenberg et al., eds., *Patterns of Racial Discrimination*, Lexington 1974, *2*, 149–66.

W. Huffman, "Decisionmaking: Role of Education," *Amer. J. of Agri. Econ.*, Feb. 1974, *56*.

F. T. Juster, "Education, Income, and Human Behavior: Introduction and Summary," in F. T. Juster, ed., *Education, Income, and Human Behavior*, New York 1975, *1*, 1–43.

N. Khaldi, "The Productive Value of Education in U.S. Agriculture," unpublished doctoral dissertation, Southern Methodist Univ. 1973.

W. Landes, "The Economics of Fair Employment Laws," *J. of Polit. Econ.*, July/Aug. 1968, *76*, 507–52.

A. Leibowitz, "Home Investments in Children," *J. of Polit. Econ.*, Mar./Apr. 1974, *82*, Supp. 111–31.

L. Lillard, "The Distribution of Earnings and Human Wealth in a Lifecycle Context," unpublished paper, Nat. Bur. of Econ. Res., New York 1974.

J. Mincer, "On-The-Job Training: Costs, Returns, and Some Implications," *J. of Polit. Econ.*, Oct. 1962, *70*, 50–79.

———, *Schooling, Experience, and Earnings*, Nat. Bur. of Econ. Res., New York 1974.

J. Morgan and M. David, "Education and Income," *Quart. J. of Econ.*, Aug. 1963, *77*, 423–37.

F. Patrick and E. Kehrberg, "Costs and Returns of Education in Five Agricultural Areas of Eastern Brazil," *Amer. J. of Agri. Econ.*, May 1973, *55*, 145–53.

S. Rosen, "Learning and Experience in the Labor Market," *J. of Human Resources*, Summer 1973, *7*, 326–42.

S. Sidhu, "The Productive Value of Education and Agricultural Development," unpublished paper, Univ. of Dar es Salaam, Tanzania, Fall 1974.

J. Smith and F. Welch, "Black-White Earnings and Employment: 1960–1970," Rand Corp., R–1666 DOL, Feb. 1975.

M. Spence, "Competitive and Optimal Responses to Signals: Analysis of Efficiency and Distribution," *J. of Econ. Theory*, 1974, *7*, 296–332.

J. E. Stiglitz, "Theories of Discrimination and Economic Policy," in G. M. Von Furstenberg et al., eds., *Patterns of Racial Discrimination*, Lexington 1974, *2*, 5–26.

P. Taubman and T. Wales, *Higher Education and Earnings*, New York 1974.

W. Vroman, "Changes in Black Workers Relative Earnings: Evidence From the 1960's," in G. M. Von Furstenberg et al., eds., *Patterns of Racial Discrimination*, Lexington 1974, *2*, 167–96.

F. Welch, "Black-White Differences in Returns to Schooling," *Amer. Econ. Rev.*, Dec. 1973, *63*, 893–907.

———, "Education and Racial Discrimination," in O. Ashenfelter and A. Rees, eds., *Discrimination in Labor Markets*, Princeton 1974a, *1*, 43–81.

———, "Education in Production," *J. Polit. Econ.*, Jan./Feb. 1970, *78*, 35–59.

———, "Relationship Between Income and Schooling," *Rev. of Res. and Educ.*, 1974b, *2*, 179–201.

K. Wolpin, "Education and Screening," unpublished paper, Univ. of Chicago, Oct. 1974.

C. Wu, "Education in Farm Production: A Theoretical and Empirical Investigation," unpublished paper, Univ. of Alabama 1972.

The Problem with Human Capital Theory—A Marxian Critique

By SAMUEL BOWLES AND HERBERT GINTIS*

Neoclassical economists have long treated labor as a commodity. They have integrated work and the worker into their analytical framework by assuming the labor-wage exchange identical to other exchanges. Because the capitalist system appears to strip the social process of work of its nonexchange characteristics, neoclassical theorists could long rest satisfied with a simple commodity interpretation of labor. And with steadily increasing levels of resources devoted to the preparation of labor, in the form of child rearing, education, health, and training, the shift toward treating the worker more precisely as a capital good seems, at least in hindsight, to have been virtually inevitable.

Human capital theory allows fundamental insights not available to earlier versions of neoclassical economics. First, it returns to and extends the Ricardian and Marxian tradition in treating labor as a produced means of production, whose characteristics depend on the total configuration of economic forces. Second, it rejects the simplistic assumption of homogeneous labor and centers attention on the differentiation of the labor force. Third, it brings basic social institutions (such as schooling and the family), previously relegated to the purely cultural and superstructural spheres, into the realm of economic analysis.

Yet this degree of success is secured at a considerable price: "labor" disappears as a fundamental explanatory category and is absorbed into a concept of capital in no way enriched to handle labor's special character. One gets the uneasy feeling that the operation was successful, but the patient vanished! The only specific attributes which "labor" retains in the human capital formulation derives from the fact, stressed by Alfred Marshall, that labor is embodied in human beings.

Human capital theory is the most recent, and perhaps ultimate, step in the elimination of class as a central economic concept. Beginning with the decline of Ricardian economics in England in the 1830's, non-Marxian economic theory has moved steadily away from attributing control of factors of production to identifiable groups and toward a theory of factor payments which self-consciously abstracts from the specific nature of the productive factors involved. In modern general equilibrium theory one can hardly tell the inputs from the outputs, much less distinguish among any specific inputs. Human capital theory is an expression of this tendency: every worker, the human capital theorists are fond of observing, is now a capitalist.

We find the theory to be substantially misleading both as a framework for empirical research and as a guide to policy. Our critique is not based on some of the commonly recognized shortcomings of the theory; for example, that market imperfections, monopoly and labor unions in particular, drive a wedge between marginal products and wages. Nor will we espouse the "training robbery" interpretation of schooling. While we recognize that educational credentials perform an important function, we are not ready to reduce the school system's economically relevant activities to screening and labeling; we believe that the evidence strongly supports the view that schooling enhances worker productivity. Nor will we dwell upon the problems which arise because skills are embodied in human beings: incomplete capital markets, imperfect information, and the possible divergence of interest between decision makers (parents) and recipients of the investment (children). Nor will we base our critique on the repugnance some have expressed at the notion of placing a dollar value on human beings. Our critique is, we believe, rather more fundamental.

---

* University of Massachusetts at Amherst. We would like to acknowledge financial support from the Ford Foundation.

By restricting its analysis to the interaction of exogenously given individual preferences, raw materials (individual abilities) and alternative production technologies, human capital theory formally excludes the relevance of class and class conflict to the explication of labor market phenomena. However, in our view such basic phenomena as the wage structure, the individual attributes valued on the labor market, and the social relations of the educational process itself can only be accounted for through an explicit class analysis. (A more detailed statement of this argument and the following material, with extensive documentation and references, is presented in our forthcoming book.)

Capitalism is a system in which the means of production are owned and controlled by a small minority. The mass of individuals, cut off from control of productive resources, are forced to sell their labor power to exist. Schooling, occupational training, child rearing, and health care perform dual economic functions: they play an essential, if indirect, role in production; and they are also essential to the perpetuation of the entire economic and social order. These processes can not be understood without reference to the social requirements for the reproduction from period to period of the capitalist class structure, the individuals who compose it, and the economic institutions that regulate it. Thus, an adequate theory of human resources must comprise both a theory of production and of social reproduction. The theory of human capital offers no theory of reproduction at all and presents a very partial theory of production, one which abstracts from the social relations of production in favor of technical relations. In the remaining sections we will argue that the failure to encompass social relations and to offer a theory of reproduction accounts for the more serious shortcomings of the standard treatments of the demand for human capital by firms, the supply of human capital, and the interpretation of the theory's central analytical concept: the rate of return to human capital.

I. The Demand for Human Capital by Firms

Our critique of human capital theory stems from a Marxian theory of the capitalist firm, one which is quite at odds with the usual neoclassical treatment.

First, we view production as a social as well as technical process. We reject the neoclassical notion of the firm as a black box whose inner workings are of interest, perhaps to the organization theorist, but not to the economist. The firm has sociopolitical dimensions which the economist may abstract from only at the cost of significant error.

Second, production is always joint production, constituting a transformation of raw materials into products and of workers with given skills and types of consciousness into workers with altered (or stabilized) skills and consciousness.

Third, labor is not a commodity, but rather an active agent whose efforts on behalf of its own objectives must be channeled, thwarted, or used to generate profits.

Fourth, the structure of wage rates is not exogenous to the firm, but rather one of the instruments used to maximize profits.

In this framework, to treat the labor-wage exchange as a pure market exchange is to abstract from an essential element in the capitalist organization: the power of the capitalist over the worker. Pure exchange relations involve no explicit power relations because the coercive instruments guaranteeing the integrity of the exchange "contract" lie outside the jurisdiction of the exchanging parties. In commodity exchange "What you see is what you get," and failing this, the aggrieved party has significant legal recourse. Not so in the case of the labor exchange which is marked by the absence of a real *quid pro quo.*

We may formalize this argument by introducing Marx's distinction between "labor" and "labor power." Labor power is the capacity of the individual to contribute to the production process in its current technical and organizational form. Thus labor power is a commodity, defined by the worker's physical and mental capacities and skills, behavioral characteristics, and potential impact on the performance of others involved in cooperative production. The labor power of the individual specifies the maximum level of performance evincible from him or her by the capitalist, given current technical and organizational conditions. The market aspect of the relation of worker to capitalist is as follows: the prospective employee agrees to surrender disposition over his or her labor power to the capitalist in return for a wage. The "labor" or actual work supplied by the individual, by contrast, is not determined in the sphere of exchange relations at all. Labor, which represents the concrete activity engaged in during the production process, depends in an essential way on the social and political structure of the enterprise.

Peak-level performance by workers benefits the capitalist rather than the worker. Actual labor productivity will normally fall short of the maximum possible with the available labor power. However, the value of the (average) product of labor will normally exceed the value of labor power (the wage), giving rise to profits. Profits on a given stock of money capital are thus maximized by exploiting the labor power of each worker to the fullest, that is, by getting as much labor as possible out of each worker. Profits are also enhanced by holding down the value of labor power, the wage. Both objectives are pursued through a judicious choice of workers and a proper specification of the technical, organizational, and political structure of the enterprise. The social organization of production is in large measure a reflection of the capitalists' need for incentive and control mechanisms which will extract labor from workers at the lowest possible wage and prevent the formation of worker coalitions which could oppose their power.

The class nature of the production process follows immediately upon observing that incentive and control mechanisms and the types of potential workers coalitions are dependent on the ownership and control structures of the enterprise. The worker attributes, which are valued by employers and which therefore constitute "human capital," are not limited to technical skills and abstract productive capacities. In particular, such ascriptive attributes as race, sex, age, ethnicity, and formal credentials, often held to be irrelevant in the logic of capitalist production, are used to fragment the work force and reduce the potential formation of coalitions within the firm. "Modes of self-presentation," to use Erving Goffman's term, such as manner of dress, speech, personal demeanor and life-style, self-concept, and status identity, can serve the same ends, while in addition insuring a relatively undistorted transmission of directives down and information up the ladder of hierarchical authority of the enterprise. Moreover, as Richard C. Edwards has shown, such work-relevant personality traits as resignation to or approval of the structure of control and distribution of rewards in the enterprise, dependability and orientation to authority within this structure, and propensity to respond individualistically to incentive mechanisms are directly relevant to extracting work from workers.

The allocation of workers to job slots, the structure of jobs available, and the definition of "productive" worker attributes simply cannot be derived, as the human capital theorists would have it, from a market-mediated matching of technically defined skills with technically defined production requirements. Issues of power, and ultimately of class, enter on a rather fundamental level.

II. The Supply of Human Capital

According to human capital theory, individuals exhibiting a particular subjective rate of time preference, and faced with an array of jobs having specified pecuniary and nonpecuniary attractions and requiring certain skills, will embark upon a course of investment in personal development. The supply of human capital is the simple aggregation of these individual choices. The demand for those services which turn raw potentials into developed capacities is derived from the individual demand for human capital. These services, "education" for short, account for the treatment of labor as a "produced means of production." In this view, the history and current state of education is the product of individual choice constrained, of course, by available educational and production technologies and the total supply of resources. Likewise, an individual's pattern of personal development (in short, how one "turns out") is depicted as the product of one's own or one's family's choices, limited only by one's "abilities," by the available learning technologies, and to a limited extent by one's family resources.

An examination of the dynamics of personal development and of the *U.S.* educational system presents a substantially different and far more complex picture. No doubt education is in part organized to produce worker traits demanded by employers. Indeed we have elsewhere sought to document the proposition that schools produce "better" workers primarily through the structural correspondence of the social relations of education with those of capitalist production, rather than through the content of the academic curriculum. Yet the social organization of schooling can in no way be depicted as the result of an aggregation of individual choices. The history of educational innovation indicates clearly that the social relations of education were rarely a reflection of popular demands, expressed either through the market or the political process. Recent research by Michael Katz, David Tyack, and others reveals that changes in classroom techniques, school finance and control, and other aspects of the organization of schools were more often than not introduced by professional elites against popular resistance.

Moreover, the production of "better workers" cannot be understood simply by reference to how individual worker-skills are related to individual worker-productivities. A highly skilled work force is not necessarily a profitable work force. Thus the educational system does much more than produce human capital. It segments the work force, forestalls the development of working class consciousness, and legitimates economic inequality by providing an open, objective, and ostensibly meritocratic mechanism for assigning individuals to unequal occupational positions. For example, the use of *IQ* and cognitive performance measures is essential to the legitimating function of schooling. Yet the use of these tests as screening devices in education goes far beyond their objective relevance to job adequacy and probably inhibits the most rational allocation of human resources. This fact is hardly compatible with the individual choice-aggregation approach and suggests that more basic social forces may be at work.

We believe these social forces derive from a basic contradiction in capitalist development: the accumulation of capital, central to the expanded reproduction of the economic system, at the same time undermines the process of reproduction of the capitalist order through the creation of a large, potentially class-conscious body of wage-workers and through dramatic shifts in the class structure. The structure of *U.S.* education, we have argued elsewhere, evolved historically in response to struggles arising from this basic contradiction. Many of these struggles have operated, in true pluralistic fashion, through the relatively uncoordinated "investment" decisions of individuals and groups as mediated by local school boards, the market for private educational services, and other decentralized decision-making arenas.

Yet the accommodation of the educational system to a changing economic reality, however pluralistic, is in essence a process led by the changing structure of production. And the evolution of the structure of production is governed by the pursuit of profit and privilege by those elements of the capitalist class which dominate the dynamic sectors of the economy. The process of individual choice aggregation, even when it is relevant to educational change, works within economic constraints determined almost entirely outside both the consumer's and the citizen's arena of choice.

Moreover, during critical periods in which basic institutional changes have been undertaken, 1840–60 or 1890–1920, for example, the individual choice model appears to be of dubious relevance to the changing structure and extent of schooling. Rather, an explicit class analysis seems best to explain the process of change and accommodation.

We are led to reject the individual choice model as the basis for a theory of the supply of educational services. The model is not wrong—individuals and families do make choices, and may even make educational choices roughly as described by the human capital theorists. We reject the individual choice framework because it is so superficial as to be virtually irrelevant to the task of understanding why we have the kind of schools and the amount of schooling that we do. It is no more able to shed light on why we "turn out" the way we do. The history of *U.S.* education or the process of individual development could probably be described in terms of functions measuring the returns to human capital and cost of capital. But what would be gained by this? The human capital approach, because it rigidly eschews class concepts, cannot offer a compelling explanation of the location of these functions or why they shift as they do, and thus cannot provide a useful theory of the supply of educational services.

We would like to offer two additional criticisms. First, economically relevant skills are not unidimensional; they cannot be aggregated into a single measure of which some individuals will have more and others less. Even in a purely formal sense, the reduction of heterogenous labor to a single "human capital" measure presents serious problems in a general equilibrium framework. Quite apart from formal problems, it is virtually impossible to make sense of the economics of human development in the United States using a model which fails to recognize that families and schools teach different things to different people—not simply more or less. The graded and cumulative academic curriculum gives the appearance of a definite hierarchy of achievement. But even in the cognitive realm we suspect that different types of learning are readily identifiable; rote-learning for some, problem-solving for others, for example. And in the noncognitive aspects of learning—the so-called hidden curriculum—different dimensions of learning are consciously either offered or imposed on children. E. B. Leacock, Jeanne Binstock, and others have shown that these differing learning contexts are closely associated with the racial, sexual, and class characteristics of the student body. Similarly, Melvin Kohn has demonstrated among families a dramatic correspondence between the objectives of child rearing and the social relations of production which confront the primary wage earner in the family.

Our second point relates to the schizophrenic usage of the word capital. The justification for the expression "human capital" is the fact that skills, like other assets, constitute a claim on future income. This usage is thoroughly in the neoclassical tradition. In the earlier, classical tradition the concept capital encompassed and unified two distinct aspects: the claim on future income and the ownership and control over the means of production. Except to the very limited extent that learning allows one to go into production on one's own, education cannot be called capital in the classical sense. Unless one accepts John Kenneth Galbraith's view of the hegemony of the technostructure (an unlikely orientation for the human capital school!), it must be admitted that educated workers do not control, much less own, the means of production. Yet it is precisely this latter, classical sense of the word capital which provides the ideological impact to the statement that every worker is now a capitalist.

III. The Returns to Schooling

Why is there a positive net return to human resource investment? What explains the pattern of rates of return? The human capital analyst, equipped with nothing more than a black box theory of both the firm and the school, is forced to offer explanations which are either superficial (supply and demand) or misleading (the interaction of tastes, technologies, and abilities). To the extent that any more specific explanations are suggested—for example, that schooling increases worker productivity through increasing individuals' cognitive capacities, and that learning abilities differ—the theory, as we have demonstrated elsewhere, is substantially incorrect. Nor are these questions unimportant. Indeed, the interpretation of the rate of return to human capital hinges upon the answer.

Our alternative theory of human resources, sketched briefly in the previous two sections, suggests the following answer to the first question, addressed specifically to the case of schooling. In the class model of the firm sketched above, the wage may under some conditions reflect marginal revenue product. But schooling will raise marginal revenue products in a variety of ways. First, investment in education may increase the labor power of the individual, either through increasing skills and productive capacities, or through supplying credentials which enhance supervisory authority. Second, schooling may increase the ease with which the employer can extract labor from a worker with given labor power by generating or selecting individual motivational patterns more compatible with the class-based power structure and incentive mechanisms of the enterprise. Third, the educated worker may be more valuable through his or her overall impact on the size of the wage bill, in that the segmentation of workers by income and status characteristics inhibits the formation of coalitions of workers capable of countering the power of the capitalist.

We may add that, because of the essential role of education in reproducing the capitalist order as a whole, the capitalist class has an interest in schooling which transcends any narrow calculation of marginal revenue products at the enterprise level. Because the capitalist class pursues its long-run interests through the state, and in important measure through its influence on educational policy, the structure of rates of return to education will reflect the often contradictory requirements of capitalist production and the reproduction of the class structure. It follows immediately that there is no reason at all to expect equality in rates of return, either among different types of schooling or between schooling and other forms of investment.

Even confining attention to the individual capitalist enterprise and thereby abstracting from class collusion by capitalists, our interpretation suggests a structure of rates of return which roughly corresponds to the observed pattern. The economic return to schooling and to age is in large part a return to a characteristic which allows the legitimate and effective exercise of authority over other workers. Black and female workers, who are by and large excluded from exercising authority over any but workers of their own sex and race, are likely for this reason to earn lower returns to schooling. Moreover, less well educated workers are, for the same reason, unlikely to gain a high return to aging; and analogously, younger workers even if white and male generally must await the arrival of a few gray hairs before enjoying any substantial return to schooling. This interpretation based on the need of the capitalist to legitimate and reproduce the power structure of the firm provides, we believe, a far more compelling explanation of the actual pattern of rates of return than does the human capital theory. The distinctly *ad hoc* and not particularly compelling attempts by human capital theorists to explain the observed pattern of rates of return is to our minds symptomatic of the limitation of the entire approach.

IV. Implications and Conclusions

Our critique of human capital theory completed, if in abbreviated form, it remains only to draw out some of the implications. We will concentrate on three major areas of application of the theory: growth accounting, distribution, and public policy.

The human capital approach to economic growth exhibits the weaknesses of the theory in a particularly straightforward manner: all factors may "contribute to growth," and barring certain technical problems, these contributions can be numerically ascertained and summed. The task of human capital researchers is thus to determine the precise contribution of investment in human resources. Even were such imputation and aggregation abstractly possible (which the now well-developed Cambridge capital critique would maintain is not), it leaves the basic questions unstated, much less answered. For we must ask not only how variations in the level of investment affect the level of output and growth rates, but also how the structure of human capital formation affects the social relations of production and the evolution of class relations. In our formulation, schooling may influence the rate of growth positively or negatively in ways which go considerably beyond the human capital theorist's notion of "labor quality": through its role in the extension and reproduction of the wage-labor system, through its capacity to attenuate class conflict and thereby to alter the rate of capital accumulation, and so on. Indeed, given that a more productive and less irrational alternative to capitalism exists, and given the role of education in reproducing the capitalist order, the contribution of schooling to growth over the last half century may on balance have been negative.

A similar narrowness is exhibited in the human capital approach to the distribution of income. Family income is the sum of returns to the various factors "owned" by the family. Thus, a measure of inequality in family income can be decomposed into dispersions in the returns to the ownership of factors and their covariances. The contribution of changes in factor endowments, say through educational equalization, to changes in the distribution of income can then be calculated. The methodological weaknesses of this approach are considerable.

First, in the human capital formulation the distribution of income is determined solely by conditions of labor supply; demand conditions are considered in a highly abstract and unenlightening manner. Macroeconomic considerations, market structure, technical change, economic dualism, and other presumably central aspects of the distribution problem are ignored. More serious, perhaps, is the naive assumption that descriptive regularities possess explanatory power. It is likely that basic regularities in the distribution of income are not directly related to human resource differences but are structural characteristics of the capitalist economy, affected by the relative power of various classes, races, sexes, and other groups. Human resource differences facilitate the assignments of individuals to places in an array of economic positions whose income structure is determined in large measure independently of the distribution of human resources.

The relationship between schooling and the distribution of income cannot be understood with a model which lacks a theory of reproduction, for a central aspect of this relationship is the role played by the school system in legitimating economic inequality. Thus, it is illogical to suppose that the reduction in inequalities in the distribution of schooling might lead to changes in income inequality in any particular direction. Major changes in the distribution of human resources will predictably be associated with changes in the structural relationships (earnings functions) relating schooling to individual income. Indeed, an equalization of education might radically reduce economic inequality, not directly, but rather by undermining the legitimacy of inequality and thus enhancing the potential for a thoroughgoing reorganization of economic institutions based on conscious class struggles or other political conflicts.

Finally, the contribution of human capital theory, through public expenditure analysis, to a balanced social policy is minimal if not actually perverse. The human capital approach, particularly when applied to schooling, embraces one of the many well-known shortcomings of public expenditure analysis: the assumption of exogenously determined individual preferences is essential to neoclassical cost/benefit analysis, yet the educational system is a prime example of an institution geared toward the alteration of preferences themselves. Given that economic returns to education are functionally related to its impact on other welfare-relevant aspects of the individual (personality, consciousness, self-concept, and interpersonal behavior), the traditional defense of cost/benefit analysis—the rigid separation of "economic" and "non-economic" welfare—bears little logical cogency.

Again, the error in the human capital approach lies in its partial view of production and its abstraction from social reproduction. For example, the repressive nature of schooling, hardly a contribution to human welfare, is an integral part of the production of a disciplined work force and is directly related to the social relations of production. Likewise, the perpetuation of sexism, racism, and elitism in our schools would not be said to be welfare-conducive or even welfare-neutral by most human capital theorists. Yet these aspects of schooling play an essential role in the reproduction of the capitalist order, a role inseparable from the capacity of schools to produce "good" workers. By abstracting from the social relations of production and the role of schooling in the reproduction of capitalism, human capital theorists have put forth a one-dimensional normative framework for the analysis of educational decisions which has no reasonable relationship to human welfare.

The theory of human capital, like the rest of neoclassical economics, ultimately locates the sources of human happiness and misery in an interaction of human nature (preferences and "abilities") with nature itself (technologies and resources). This framework provides an elegant apology for almost any pattern of oppression or inequality (under capitalism, state socialism, or whatever), for it ultimately attributes social or personal ills either to the shortcomings of individuals or the unavoidable technical requisites of production. It provides, in short, a good ideology for the defense of the status quo. But it is a poor science for understanding either the workings of the capitalist economy or the way towards an economic order more conducive to human happiness.

REFERENCES

J. Binstock, "Survival in the American College Industry," Brandeis Univ., unpublished doctoral dissertation 1970.

S. Bowles and H. Gintis, *Education and Capitalism in the U.S.*, New York 1975.

R. C. Edwards, "Individual Traits and Organizational Incentives: What Makes a 'Good' Worker?," in *Education and Psychological Measurement*, forthcoming 1975.

M. Katz, *The Irony of Early School Reform*, Cambridge, Mass. 1968.

M. Kohn, *Class and Conformity*, Homewood 1969.

E. B. Leacock, *Teaching and Learning in City Schools*, New York 1969.

D. Tyack, *The One Best System: A History of American Urban Education*, Cambridge, Mass. 1974.

SUPPLY AND MOBILITY OF WOMEN ECONOMISTS

Job Search and Affirmative Action

By ALICE H. AMSDEN AND COLLETTE MOSER*

In the first part of this paper, the recruitment and hiring practices in the job market for men and women economists in 1973–74 are examined with particular emphasis on the relationship between interviews and follow-up job offers. In the second part, the perceived effect of affirmative action implementation on employment and employment conditions over the past three years is reported.

I. Job Search in an Affirmative Action Context[1]

The 1973–74 job market for economists operated under the pressures of affirmative action (*AA*) which had been building up for a number of years. Several hypotheses about the possible effects of *AA* on the 1973–74 job market for economists were articulated by observers when recruitment and hiring were done. First, some observers suggested that, as designed and enforced, *AA* was window dressing. It was ineffective as a means to end what was assumed to be sex discrimination in employment. It also had the undesirable side effect of raising women's job search costs. Job search costs rose because prospective employers enthusiastically interviewed women without any intention of hiring them. This gave the appearance of complying with *AA* requirements. Second, some observers suggested that *AA* forced prospective employers not practicing discrimination zealously to recruit and hire women in order to be above suspicion. As a consequence of excessive interviewing, job recruitment costs rose. As a consequence of the total demand for top-rate women economists outstripping the supply, the quality of faculty declined. Third, some observers suggested that *AA* discriminated against men. Male economists had fewer opportunities than women to be interviewed and had fewer follow-up job offers.

Preliminary data from The Committee on the Status of Women in the Economics Profession (*CSWEP*), American Economic Association (*AEA*) 1974–75 Survey of Economists provide some insights into the recruitment and hiring practices of men and women economists in 1973–74. The extent to which *AA* influences such practices is as yet indeterminate and awaits time series analysis with precise quantitative measures of *AA*. This paper summarizes the information provided by those economists surveyed who participated in the 1973–74 job market.[2] It concentrates

---

* Lecturer, department of economics, University of California, Los Angeles, and assistant professor of agricultural economics, Michigan State University, respectively.

[1] Part I was written by Amsden.

[2] Information on the sampling and survey procedure and quality of sample may be obtained from *CSWEP, AEA*. Preliminary data from the 1974–75 *CSWEP, AEA* Survey of Economists used in this article are based on the 710 questionnaires returned by 55 percent of the women economists registered with *CSWEP, AEA*; the first 264 male economists in matched cohorts; and a special male job-market sample. The education cohort was selected at random from three names supplied by each woman economist of men who were in her class when the woman did her highest level of academic work. A name for the employment cohort was supplied by each woman, using an alphabetic criterion, of the male economist at her institution who holds the same rank or approximate level of job responsibility. The male job-market sample consists of two types of men thought to have been in the 1973–74 job market: (1) men whose names were supplied by women answering questionnaires (approximately 62 percent of the male job-market sample), and (2) men whose names were selected for sampling from the files of vitas received by

on the relationship between interviews and follow-up job offers of respondents. This is to shed light on the speculations which the 1973–74 job market generated. If it is found that the ratio of interviews to follow-up job offers differs between men and women when quality is controlled for, then the hypothesis that sex discrimination exists cannot be ruled out.

A distinction is drawn between three job markets. One concerns recruitment at the *AEA* meetings in New York City in December 1973, job offers directly resulting from initial interviews at the meetings, follow-up "intensive" interviews resulting from initial interviews at the meetings, and subsequent hiring. Another market concerns intensive interviews and follow-up job offers not arranged at the meetings. A third market concerns interviews and follow-up job offers unsolicited by candidates and arranged at the initiative of employers. The ratio, R, chosen to measure the relationship between initial interviews, I, and follow-up interviews or job offers, F, is the following:

$$R = (I - F)/I \tag{1}$$

The ratio can take a value of between zero and one. A high ratio indicates a poor follow-up record. Hence, R can be thought of as a measure of "excess interviewing." A value of zero indicates a perfect follow-up record.

The mean ratios calculated for job market respondents are presented in Table 1. They are broken down by state universities and colleges, private universities and colleges, government and business. Aggregate mean ratios for all four categories are also presented.

The comparative magnitudes of the five ratios for both sexes for the aggregated four categories of employers are consistent with a priori reasoning. R_1, the ratio of *AEA* initial interviews to direct follow-up job offers, is high for the aggregated four categories. As one would expect, not many employers make job offers solely on the basis of interviews at the meetings. R_5, the ratio of unsolicited job interviews to job offers, is low. Presumably interviews arranged specifically at the behest of employers are more likely to result in hiring than are all classes of interviews. It is noteworthy that, both for men and women, recruitment not initiated at the meetings, R_4, is more likely to result in hiring than recruitment initiated at the meetings.

What also emerges from Table 1 is the fairly small difference in the ratios of men and women for the aggregated four categories of employers. R_1, R_4, and R_5 are almost identical for both sexes. In the *AEA* job market, however, women tend to have a slightly better chance for a follow-up intensive interview, R_2, than do men. Men tend to have a slightly better chance for a follow-up job offer, R_3, than women. But in neither of these cases are the means for men and women significantly different from each other at the 96 percent level of confidence. The evidence suggests that, even in the year when affirmative action gained unprecedented publicity, the interview/follow-up record of women economists was no better than that of men.

When the ratios for men and women are disaggregated according to category of employer, three out of five of the ratios for private universities and colleges indicate a better follow-up record for men. Four out of five of the ratios for business indicate a better follow-up record for women. The samples from which most of the disaggre-

---

several economics department chairmen. These files were sent to the chairmen by placement officers throughout the country. Both parts of the job-market sample were relatively small (despite a fairly high response rate) and were nonrandom.

This article is based on the 262 women and 154 men who were in the job market; i.e., they had one or more job interviews or offers after October 1, 1973.

TABLE 1—MEAN RATIOS OF EXCESS INTERVIEWING FOR FEMALES (F) AND MALES (M), JOB MARKET SAMPLE, OCT. 1973–NOV. 1974

	R_1[a]		R_2[b]		R_3[c]		R_4[d]		R_5[e]	
	F	M	F	M	F	M	F	M	F	M
Mean for state universities or colleges	.99	.91	.68	.86	.67	.48	.39	.28	.32	.41
Number of respondents	39	71	37	54	25	33	63	40	42	38
Mean for private universities or colleges	.93	.97	.77	.79	.68	.54	.41	.28	.46	.32
Number of respondents	34	60	29	53	14	30	42	34	30	29
Mean for government	.86	.87	.73	.67	.50	.52	.37	.60	.47	.45
Number of respondents	23	45	16	36	11	23	48	25	32	19
Mean for business	.87	.94	.58	.88	.58	—	.45	.56	.32	.27
Number of respondents	16	29	13	27	11	7	54	21	35	16
Mean for all four categories	.92	.90	.67	.76	.62	.51	.39	.35	.37	.32
Standard deviation	.21	.21	.32	.33	.42	.45	.41	.52	.43	.42
Number of respondents	49	84	49	84	38	58	153	87	117	73

Source: Preliminary data from *CSWEP*, *AEA* 1974–75 Survey of Economists.

[a] $R_1=(m-n)/m$ where m = number of interviews at 1973 *AEA* meetings and n = number of job offers resulting from 'm'.

[b] $R_2=(m-o)/m$ where o = number of follow-up intensive interviews resulting from 'm'. By an intensive interview is meant a visit to a college, etc., for the purpose of meeting staff, presenting a paper and/or discussing employment.

[c] $R_3=(o-p)/o$ where p = number of job offers resulting from 'o'.

[d] $R_4=(q-r)/q$ where q = number of intensive interviews without prior *AEA* interview and r = number of job offers as a result of 'q'.

[e] $R_5=(s-t)/s$ where s = number of unsolicited intensive interviews and t = number of job offers resulting from 's'

gated ratios are calculated, however, are very small. Henceforth, the ratios presented will be limited to the aggregated four categories of employers.

An attempt was made to compare the interview/job-offer ratios for men and women drawn from a homogeneous sample. To control for quality, male and female respondents have been grouped according to the institution in which they undertook their Ph.D. training. This criterion for controlling quality, however, is far from perfect. It assumes a homogeneity of quality within each group that may not be warranted. The abilities of men and women with Ph.D.s from the top seven schools in economics, for example, may differ. This may be a consequence of the imposition of more stringent entrance requirements for women than for men because of a sex bias in graduate admissions policies. This would imply a higher quality of women graduates. Such a limitation must be kept in mind when interpreting the interview/follow-up ratios for men and women controlled by Ph.D. institutions. These ratios are presented in Table 2.

The evidence in Table 2 suggests that women graduates of the top seven schools in economics have a better follow-up record than men graduates of the same schools in three out of four cases. But the differences in mean ratios are not significant at the 96 percent level of confidence. The ratios for men and women graduates of other schools in the "Chairman's Group"

TABLE 2—MEAN RATIOS OF EXCESS INTERVIEWING AND NUMBER OF RESPONDENTS BROKEN DOWN BY INSTITUTION WHERE PH.D. WORK UNDERTAKEN

Item	R_2[a]		R_3[a]		R_4[a]		R_5[a]	
	Female	Male	Female	Male	Female	Male	Female	Male
Mean for top seven schools in economics	.53	.78	.64	—	.26	.38	.32	.44
Number of respondents	11	12	10	6	20	16	19	17
Mean for other schools in "Chairman's Group"	.71	.71	.58	.54	.41	.39	.40	.32
Number of respondents	29	50	20	39	61	45	57	39
Mean for other schools	—	.88	—	.34	.46	.31	.35	.27
Number of respondents	9	20	7	12	37	18	20	13

Source: Preliminary data from *CSWEP*, *AEA* 1974–75 Survey of Economists.

[a] See footnotes to Table 1 for definition.

are very similar. Male graduates of colleges and universities which are generally ranked lower academically than colleges and universities in the Chairman's Group have a better follow-up record than comparable female graduates. This may be some indication that employers have not succumbed to the pressures of *AA* by "overhiring" women with less than impeccable credentials compared to men of the same calibre. Firmer indication that academic standards have been maintained, however, would require comparing the academic rank of the institutions hiring such men with that of the institutions hiring such women.

Although men and women economists have similar interview/follow-up ratios, their opportunities for employment may differ. One indication that unequal opportunities for employment exist would be differences in absolute number of interviews for men and women. Table 3 presents the average number of initial *AEA* interviews, non-*AEA* interviews, and unsolicited interviews of male and female respondents. Such averages have also been broken down by the institution where Ph.D. work was undertaken. For each indicator of job opportunity (for example, non-*AEA* interviews), only respondents with at least one interview were sampled. This was to insure that the sample contained only respondents who were actively seeking employment. Not unexpectedly, both men and women had more *AEA* interviews on average than other types of interviews. Presumably the marginal cost of an additional interview at the meetings is less for both interviewer and interviewee than in other situations. There appears to be little indication that one sex had greater opportunities for employment than the other. This is true both when quality is controlled for and when it is not. The average number of interviews for men and women is quite similar. The possibility remains, however, that one sex found it easier than another to arrange even one interview. There may be more men than women who would have liked to be interviewed but were not.

An attempt was made to get some idea of the salaries offered to men and women job-searchers. Respondents were asked to specify both the salary of their first job offer and the salary of the job they accepted. They were also asked to break down these figures by state and private universities and colleges, government, and business. To make meaningful comparisons of male and female salary offers, however, it is necessary to control for two factors:

TABLE 3—AVERAGE NUMBER OF INTERVIEWS OF JOB-MARKET RESPONDENTS, OCT. 1973–NOV. 1974[a]

Item	Initial Interviews At The *AEA* Meetings		Non-*AEA* Interviews		Unsolicited Interviews	
	Female	Male	Female	Male	Female	Male
All universities and colleges						
Average number of interviews	7.7	8.8	2.2	1.8	1.9	2.1
Number of respondents	49	84	153	87	117	73
Top 7 schools in economics						
Average number of interviews	9.5	5.8	2.2	1.9	3.0	2.4
Number of respondents	11	12	20	16	19	17
Other schools in Chairman's Group						
Average number of interviews	8.3	9.4	2.2	1.7	1.7	1.9
Number of respondents	29	50	61	45	57	39
All other schools						
Average number of interviews	—	9.7	2.2	1.8	1.7	
Number of respondents	9	20	37	18	21	13

Source: Preliminary data from *CSWEP*, *AEA* 1974–75 Survey of Economists.

[a] Not every respondent provided information on the school from which Ph.D. was received. Therefore, the total number of respondents for "All universities and colleges" may in each category of interview exceed the subschool total.

the quality and experience of the job applicant and the character of the hiring institution. Salaries for the same position may vary by geographic region, for example. They may also vary by type of academic institution. It appears that, when academic institutions are ranked by quality along a spectrum, salaries for assistant professors approximate a normal curve. The highest and lowest ranking institutions pay the least. Differences such as these could not be controlled for on the basis of the questions respondents were asked. If, however, it is assumed that differences in institutions hiring women and in institutions hiring men are distributed randomly, the figures presented in Table 4 give a very rough indication of the comparable magnitude of the salaries offered to men and women. Only the salaries of jobs accepted are presented for all job respondents, irrespective of position. The profile of positions of male and female respondents who provided information on salaries, however, appears to be fairly similar. The figures indicate no consistent pattern of differences between the sexes. This was also the case when candidates were grouped according to institution where they undertook their Ph.D. work. Nor was there a consistent

TABLE 4—SALARIES OF JOBS ACCEPTED BY JOB-MARKET RESPONDENTS

Item	Mean Salaries of All Respondents	Number of Respondents
State universities and colleges		
Female	$13,850	35
Male	14,650	38
Private universities and colleges		
Female	14,700	26
Male	13,500	25
Government		
Female	16,600	29
Male	16,850	16
Business		
Female	17,000	24
Male	—	6

Source: Preliminary data from *CSWEP*, *AEA* 1974–75 Survey of Economists.

pattern of differences in the salaries of first job offers made to men and women.

The sample results may be summarized as follows: first, the hypothesis that the job-search costs of women in 1973–74 were higher than those of men because employers interviewed women without any intention of hiring them is not supported by the data. On the other hand, even in the presence of an intensification of *AA*, the interview/follow-up record of women economists was no better than that of men. Second, there is some indication that the quality of faculty did not decline in the presence of *AA*. Women Ph.D. graduates of relatively less prestigious schools did not have a better follow-up record than men graduates of the same academic institutions. Third, it does not appear that one sex had more opportunities for employment than the other. The average number of interviews for men and women with at least one interview is fairly similar. Fourth, the salaries offered to men and women show no consistent pattern of differences.

There is room for much more work to better explain variations in the five interview/follow-up ratios. Additional coding of the sample data would permit regressions to be run for each of the five dependent ratios. One set of explanatory variables might include marital status, children, and mobility of spouse. Such variables are believed to influence the decision to hire women in particular. Another set of independent variables to explain the relationship between recruitment and hiring might include institution where Ph.D. work was undertaken, prime field of specialization, academic position, helpfulness and prominence of thesis adviser, and number of publications. Sample data on all these variables have been collected. This type of analysis, however, is only one of the many directions which further research might take.

II. Perceptions of Affirmative Action Issues[3]

The late 1960's and early 1970's witnessed the implementation of several pieces of legislation relating to the employment of women. Notable among these are the Equal Pay Act which requires that men and women doing substantially the same work receive equal pay; Title VII of the Civil Rights Act which mandates the removal of various discriminatory employment activities; amendments to the Higher Education Act and Executive Order 11246 and Revised Orders 4 and 14 requiring that employers with government contracts develop affirmative action plans for improving women's employment opportunities and equity.

Affirmative action policies relate to the education and employment of women economists as well as women in other occupations. Some areas for affirmative action are as follows: recruitment and interviewing; the numbers of women hired in previous "male only" jobs; job assignments and responsibilities; wages, fringe benefits, and other methods of remuneration; promotion and upward mobility; treatment with respect to layoffs, discharges, and hours of work.

Implementation of affirmative action policies involves both direct changes in the variables cited in the previous paragraph and indirect personal interaction changes which might affect one's perception of the "acceptance" of a woman in an occupation. Such acceptance is expected to generate an increase in the supply of women in that field.

Given these affirmative action principles and policies, the Committee on the Status of Women in the Economics Profession included in its survey a set of questions with the following major purposes:

---

[3] Part II was written by Moser.

(1) to evaluate economists' perceptions of changes in areas of employment related to affirmative action implementation, and

(2) to evaluate economists' perceptions of the continuation of differential treatment based on sex even within this affirmative action implementation period.

The affirmative action question set, Part IV of the survey, included seven separate questions. One of these questions was a "grid" of ten areas of employment or employment conditions in which one might expect some perceived change as a result of affirmative action implementation. The phrase "during the last three years" was used as a proxy for the implementation period. Sample size was 710 women and 332 men. The response rate for most of the affirmative action questions was 95–98 percent of this sample. On the ten-question "grid," the response rate fell to an average of about 65 percent for women and 55 percent for men.

The majority of the respondents were situated in establishments which had affirmative action plans for women. Fifty-seven to 60 percent of all men and women respondents said that their establishment had either a written or verbal plan. Only 4 percent of the women and 7 percent of the men said that their establishment *did not* have such a plan.

Answers to the ten-question grid concerning areas of employment conditions within the last three years are particularly relevant given this degree of awareness of existence of institutional affirmative action plans. Respondents were asked whether they had observed certain employment changes with respect to themselves (if they were women), and with respect to other women economists; whether no change had been observed; whether the issue really was not "a problem"; or whether opportunities for women economists had actually worsened in this area. Since responses to the last outcome (opportunities worsened) were minimal, they were dropped from the reported tabulations. Responses to the first two outcomes were combined to form a category which indicated improvements for women.

Six of the affirmative action employment-related issues lent themselves to the above kinds of treatment. These areas are as follows: (1) increased access to tenure, seniority, or promotion for women; (2) affirmative action salary adjustments for women economists; (3) increased administrative responsibilities for women, including increased appointments to important committees; (4) increased support of women economists by male colleagues; (5) increased support of women economists by students and other clients; (6) increases in fringe benefits and such perquisites as financial support for conference attendance.

The most frequent observation on each of these questions related to improvements in employment conditions for women economists was that no change had occurred. Only in the area of increased administrative responsibility and committee assignments did the number of women respondents indicating change almost equal the number observing no change. Women saw the least amount of change in the area of fringe benefits.

Another consistent outcome for each of these employment issues was that one and a half to twice the percentage of men to the percentage of women *did not* see the issue as a problem for women economists in their department or division. Table 5 summarizes these results.

In an effort to determine the extent of perceived employment discrimination by sex during this three-year period of affir-

TABLE 5—PERCEPTIONS OF AFFIRMATIVE ACTION AREAS OF EMPLOYMENT FOR WOMEN ECONOMISTS

Observations	Issues					
	Tenure and Promotion	Salary Adjustments	Administrative Responsibility	Male Colleague Support	Client Support	Fringe Benefits
	(All Results are in Percentages)					
Women observed change	33.2	25.7	40.7	24.7	29.6	10.3
Men observed change	33.9	27.2	31.6	24.3	23.8	4.3
Women observed no change	45.6	51.8	41.5	54.5	48.0	66.7
Men observed no change	37.2	43.2	41.2	48.6	43.4	61.3
Women: "not a problem"	19.0	20.6	16.9	19.2	22.4	21.5
Men: "not a problem"	35.0	34.9	33.9	34.1	35.1	41.1

Source: Preliminary data from *CSWEP*, *AEA* 1974–75 Survey of Economists.

mative action implementation, respondents were asked a few other questions related to employment conditions, employer preferences, and interviewing tactics.

A key question was whether respondents working as economists thought that they were paid less or had a lower job level than what they would have if they were of the opposite sex. Of the women respondents, 34.7 percent believed this to be the case. A much smaller percent of the men, 13, believed they were paid less or had a lower position because they were men. In the open-ended part of this question which asked "yes" respondents to indicate how this "gap" occurred, a sample of male respondents overwhelmingly asserted affirmative action to be the cause. Women respondents gave much more mixed interpretations. Of a sample of about 300 female respondents, the most frequent explanation given was that the gap occurred at hiring, oftentimes because the woman was not an aggressive bargainer or through her lack of comparative salary data.

Very frequently women economists cited restrictions on promotion and limited access to administrative positions as the cause of the earnings and position gap. Attribution of the gap to discontinuous labor force association, a factor emphasized by G. E. Johnson and F. P. Stafford was given by only four of the women respondents in the sample.

Respondents were also asked whether in the last three years an employer had indicated in their presence a preference for male rather than female economists. About 23 percent of the women and 9 percent of the men indicated that this situation had occurred. The magnitude of the positive response to this question is surprising in this affirmative action period. Also inconsistent with affirmative action principles were the "reasons" which respondents said employers gave for their preferences. While "children and family responsibilities" were cited by several respondents, subjective perceptions of female characteristics, such as "unstable," "not serious," "not smart," "can't get along with coworkers," were given to a larger number of female and male respondents as the employer's explanation for preferring male to female economists.

While the previous question did not necessarily refer to employer preferences for men expressed at the time of an interview, a question was asked on an interviewing procedure wherein the employer's questioning concentrates on nonprofessional competency issues. Of the women who had taken interviews in the last three years, 160 or 31.9 percent reported that they had received interviews in which a disproportionate amount of questioning was related to personal and domestic issues. Only 11 men (4.3 percent of the males taking interviews) reported that they had been subjected to this kind of interviewing.

The early 1970's witnessed large-scale attempts to implement affirmative action employment concepts. Responses to the *CSWEP* survey affirmative action questions indicate that about one-fourth to one-third of our respondents are observing improvements for women economists in terms of their conditions of employment. Analysis of other questions related to perceived acts of employment discrimination against women indicates that a significant portion, perhaps one-third, of our women respondents feel that they have been subjected to this kind of discriminatory treatment even during this assumed period of affirmative action implementation. The strength of the no-change responses, as well as the indications of continued differential and unfavorable treatment of women economists, does not suggest a reversal of attitudes. Even with affirmative action programs, improvements for women economists seem to be marginal.

REFERENCE

G. E. Johnson and F. P. Stafford, "The Earnings and Promotion of Women Faculty," *Amer. Econ. Rev.*, Dec. 1974, *64*, 888–903.

Women Economists: Career Aspirations, Education, and Training

By MYRA H. STROBER*

The economics profession is, at the present time, predominantly male. According to the 1970 census, women are only 12 percent of all economists and only 14 percent of economists teaching at colleges and universities. Moreover, among those currently training to become economists, the percentage of women is also quite low. In a 1973 study of the fifty-seven major American economics departments (see Carolyn Bell, 1974) the American Economic Association (*AEA*) Committee on The Status of Women in the Economics Profession (*CSWEP*) found that women are only 12 percent of current full-time Ph.D. candidates, 14 percent of full-time M.A. candidates, and 15 percent of senior economics majors.

Occupational segregation, such as that existing in the economics profession, is beginning to be recognized as a serious misallocation of resources; and efforts to reduce such segregation are now being made in many places. At its December 1971 meeting, the *AEA* resolved to "redress the present low representation of women in the economics profession." (See American Economic Association, 1972, p. 447.) Clearly, there are both demand- and supply-side factors operative in occupational segregation. However, this paper deals only with the supply side. For in our view, to successfully reduce occupational segregation, it is critical to understand the supply-side factors which induce women to seek careers in economics and assist them in completing their graduate training. The results described here are based on the 1974–75 *CSWEP*, *AEA* Survey of Economists and should be considered preliminary as data collection is still underway.[1]

I. Career Aspirations

When and why do women decide to become economists? About three-fourths (74 percent) of the women in our sample made the decision to become an economist during college. Only 3 percent made that decision before they entered college and about 24 percent after college (13 percent during graduate school and 11 during post-college employment). These patterns contrast with those in Helen Astin's study, which found that one-fourth of the women doctoral recipients for 1957–58 knew what field they would specialize in even before they completed high school; with a study by Rita Stafford of outstanding women in New York state, which found that 26 percent of the lawyers and 45 percent of the physicians had made their career decisions before they entered college; and with a study by Erwin Linn, which found that 47 percent of women dentists who were graduates of *U.S.* dental schools had decided to become a dentist before entering college. On the other hand, over one-third of the women in Astin's sample said they had decided on their field after college graduation or during graduate school. It would appear that the timing of the decision by a woman to enter a predominantly male field is quite variable and may be related to public familiarity with the particular occupation.[2] Opportunities to influence career decisions seem to be present during a long period of a young woman's life. It is interesting, too, that, although most women economists waited until college to make a *specific* career choice, almost two-thirds (60 percent) had decided while still in high school that they would have a career. And of these, 50 percent planned to pursue nontraditional careers.[3]

---

* Assistant professor of economics, Graduate School of Business, Stanford University. I should like to thank Chiranjib Sen for programming assistance and Patti Osborne for help with coding.

[1] Information on the sampling and survey procedure and quality of sample may be obtained from the author. Preliminary data used in this article are based on 678 of the 710 questionnaires returned by 55 percent of the women economists registered with *CSWEP*, *AEA*. (Thirty-two women who indicated they had no primary field in economics were excluded.) Each woman economist was requested to supply the names of three male economists who were in her class when the woman did her highest level of academic work in economics. From those names supplied, one name was chosen at random. There are 130 men in the matched sample. The numbers of men and women in the matched sample who have Ph.D.s are unequal because some of the men, matched with women at the bachelor's or master's level, later went on to receive Ph.D.s.

Which persons and factors influenced women's career aspirations during their high-school years? Seventy percent of the women in our sample indicated that they had at least one important role-model in high school. Half of the women had at least two such persons and about one-third had three. For those who had at least one role-model, the most frequently mentioned were a parent (44 percent), or a teacher, administrator, or employer (35 percent). Other relatives and friends were less frequently mentioned role-models.

In about 49 percent of the cases, the most important high-school role-models were men; in about 46 percent women; and in 5 percent both. Where a teacher, administrator, or employer was the most important role-model, the individual was a man in about half (53 percent) of the cases. However, where a parent was the most important role-model, it was the woman's *father* who was the most important influence in 81 percent of the cases. Thus, of those women economists who had an important role-model during high school, only 6 percent saw their career plans as primarily influenced by their mothers. And this was despite the fact that 42 percent of our economists had working mothers who were highly educated and whom they wished to emulate. (Of those whose mothers were employed, 60 percent said they saw their mothers' experience in working outside the home as something they wished to emulate.) Perhaps the fact that these mothers were largely (62 percent) in traditional occupations made them less influential for their daughters.[4]

The ability of the majority of women economists' families to finance their daughters' aspirations, at least through college, should not be minimized. Women economists have relatively highly educated fathers who, presumably, earned relatively high incomes. In our matched sample, there is a significant difference (at the 3 percent level) in education levels between the fathers of men and women economists.[5] A greater percentage of the fathers of women economists are college graduates (17 percent for the women as against 13 percent for the men), have some graduate school training (9 percent versus 1.5 percent), and hold a graduate degree (29 percent versus 23 percent). Also there is a significant difference (at the 1 percent level) in education levels between the mothers of women and men economists. A greater percentage of the mothers of women economists have some college training (30 percent for the women versus 17 percent for the men), are college graduates (19 percent versus 11 percent), and hold graduate degrees (14 percent versus 9 percent). In his study of science doctorates, Lindsey Harmon also found that the fathers of women doctorates were more highly educated than those of men doctorates, and John Folger, Astin and Alan Bayer, in their study utilizing the Project Talent Data Bank, found that "socioeconomic background is more important for girls than for boys in determining college completion and entry to a professional occupation" (p. 287).

---

[2] Only 5 percent of the men in the matched sample made the decision to become an economist before they entered college.

[3] Accountant, mathematician; engineer, architect, physical or life scientist; physician, dentist, pharmacist; social scientist; college teacher; lawyer, politician; manager or administrator; entrepreneur.

[4] Traditional occupations are: librarian; nurse, dietician, home economist; therapist, health technologist; social or recreation worker; nursery, elementary, or high school teacher; sales worker; clerical worker.

[5] A Chi-square test was used to calculate differences between male and female economists.

For those women who had role-models in high school, the most important lessons they seem to have learned from their "mentors" were: that a career can be interesting and fulfilling (87 percent); to seek a high degree (63 percent); and to seek a highly paid job (58 percent). One set of skills they indicated they did *not* frequently learn was how to manage simultaneous family and career responsibilities. And indeed, given our societal mores, perceived potential difficulties or actual difficulties with dual responsibilities may be one reason why, in our matched sample, the percentages of women who are single (26 percent) and divorced or separated (12 percent) are so much higher than the percentages of men (12 and 6 percent, respectively).

As noted earlier, 77 percent of our sample had decided to be economists by the time they had graduated from college. And, indeed, 67 percent of the women majored in economics in college. Clearly, influences during the college years are central to explaining women's decisions to pursue an economics career. Which persons and factors influenced women's career decisions during their college years?

Eighty percent of the women in the sample said there was at least one person during their college years who exerted an important influence on their decision to pursue a professional career. Of those who had an important influence, it was most often (56 percent) a professor, employer, or college/university counselor; frequently (25 percent) a parent; and less often a relative, friend, spouse, or fiance. Sixty-one percent of the most important college influences were men.

While in college, about 30 percent of the women knew a person who tried to *dissuade* them from pursuing a professional career. In 40 percent of the cases, the dissuader was a parent, but teachers and employers accounted for 20 percent of the dissuaders, near relatives for 17 percent, and friends for 13 percent. Less often (7 percent) fiancés or boyfriends, or husbands (4 percent) acted as dissuaders. In 51 percent of the cases, the dissuaders were men; in 43 percent, women; and in 7 percent, both. Only 10 percent of the men in the matched sample knew someone in college who tried to dissuade them from pursuing a professional career.

Why, specifically, did the women in our sample choose to become economists? At the time they decided they wanted to be economists, 45 percent of the women in our sample said they made this decision because they liked the subject matter of the discipline (i.e., they "found it interesting" or "liked economics courses"). An additional 14 percent were attracted to economics because they perceived that they could use the subject matter to help people or solve social problems, and 7 percent chose to be economists in order to utilize their interest in mathematics. However, despite the fact that they chose to deal professionally with economic concerns, only 14 percent of the women in the sample chose to be economists primarily because of economic factors, i.e., attractive salaries or job opportunities. An additional

9 percent of the women mentioned attractive salary or job opportunities as a second important reason for becoming an economist. The ability to solve social problems or help people was mentioned as a second important reason for becoming an economist by 9 percent of the women and ability to apply mathematics as a second reason by 5.5 percent.

In the matched sample, we find no significant differences between men and women with respect to their motivations for choosing to become economists. Forty percent of the women and 36 percent of the men became economists because they liked the subject matter. However, the men more frequently mentioned job and salary opportunities as their most important reason for choosing economics (15 percent) than did the women (8 percent) and the women more frequently mentioned "to solve social problems" and "help people" (18 percent) than did the men (15 percent).

Although the ability to utilize mathematics was not often a major reason cited for choosing a career in economics, nonetheless the women in our sample had a strong mathematics background. In high school their mean number of years of mathematics was 3.4. This is an important finding, for it supports those who observe that the absence of a strong mathematics background can inhibit women who might otherwise choose to major in economics. In college, too, many women economists studied a good deal of mathematics. About 80 percent studied statistics, 63 percent calculus, 30 percent matrix algebra, and 30 percent "advanced" mathematics. In the matched sample, there were no significant differences between the percentages of women and men who studied these four subjects.

What types of colleges and universities did the women in our sample attend? The majority of women (58 percent) went to private colleges or universities, and most (75 percent) attended a coeducational school (either private or public). However, a substantial minority (26 percent) attended a single-sex school. This finding is noteworthy, as there is considerable literature on the value of single-sex schools in influencing women to pursue professional careers. (See, for example, Elizabeth Tidball.) Also, recently, there has been some evidence that women are more likely to major in economics at an all-girls college than in a coeducational setting (Bell 1974).

If we classify institutions by quality, we find that only a very small proportion of women undergraduates (8 percent) attended a school where the economics department was one of the top seven in the country.[6] In part, this is because several of the undergraduate schools with which the top seven departments are associated are, or have been, all male. Twenty-one percent of those in the sample went to schools in the so-called Chairman's Group or to a top-ranked British or French institution; 17 percent attended a top-ranked small, undergraduate college; and 12 percent went to schools at which the American Association of University Professors gave its highest or second-highest ranking for full professors' salaries. Thirty-eight percent went to none of these types of schools but to some other institution.

---

[6] Institutions were divided into five categories. The first category consists of the seven schools which, according to the Cartter Report, have the top economics departments in the country: M.I.T., Harvard, Yale, Stanford, University of California at Berkeley, Princeton, and University of Chicago. The second category consists of schools in the so-called "Chairman's Group," (see Bell 1973) except for the seven included in the first category. Oxford, Cambridge, The London School of Economics, and the University of Paris are also included in category two. Thus, there is a total of forty schools in this category. In the third category are thirteen small undergraduate colleges to which in 1973 the American Association of University Professors (*AAUP*) gave its highest or second-highest (1* or 1) ranking for full professors' salaries: Amherst, Antioch, Barnard, Goucher, Grinnell, Oberlin, Pomona, Smith, Swarthmore, Vassar, Wellesley, Wesleyan, and Williams. The fourth category consists of 132 other schools to which the *AAUP* in 1973 gave its highest or second-highest ranking for full professors' salaries. In the fifth category are all American and foreign schools not in one of the first four categories.

In looking back on their college careers, about 21 percent of the women felt that a narrow perception of their future employment opportunities limited their choice of undergraduate college. Thirty percent felt that a narrow perception had limited their undergraduate course work. However, some men, too, had these feelings, and in the matched sample there were no significant differences between the proportions of men and women who felt that narrow perceptions had limited their choices.

What specialties within the field of economics do women enter? The evidence at hand thus far indicates that there are some differences between the distributions of men and women among the specialties in economics. In comparing the percentage of women in the ten economics specialty fields with the percentage of all *AEA* members (male and female) in these fields, we found a significant difference (at the 1 percent level) in the distribution among fields.[7] A higher percentage of women in our sample were in the fields of labor (19 percent); welfare programs, consumer economics, and urban and regional economics (12 percent); and general economics, theory, and history (19 percent), as compared to the *AEA* membership as a whole (10, 8, and 14 percent, respectively). In the matched sample, among *AEA* members only, there was also a significant difference (at the 4 percent level) in the proportions of women and men in the ten fields. Again, a higher proportion of all women were in the fields of labor (25 percent); and welfare programs, consumer economics, and urban and regional economics (16 percent), as compared to men (9 and 10 percent, respectively). However, this time men were more heavily represented in the field of general economics, theory, and history (26 percent) as compared to women (15 percent). Clearly, the evidence is conflicting with respect to general economics, theory, and history. However, it does appear that women economists are more likely than their male counterparts to specialize in the two fields of labor and welfare programs, etc. These choices may be partly related to the slight, though statistically insignificant, differences in motivations to study economics described above. It should be pointed out, however, that the tendency toward some inequality in distributions among subfields is not nearly as marked in economics as it is in other predominantly male fields, for example medicine and law (see Michelle Patterson).

II. Education and Training of Women Ph.D.s in Economics

In this section we explore three issues in the education and training of women Ph.D.s: financing of degree, interruption of training, and attrition. How do women finance their graduate training? For the total female Ph.D. sample ($N=300$) and for both the men ($N=86$) and women ($N=72$) in the matched Ph.D. sample, fellowships at the predissertation as well as the dissertation stages were a major source of funds for 50–70 percent of our respondents, and in the matched sample, for both stages of study, there were no significant differences between the proportions of men and women who said that fellowship aid was a major or minor source, or not a source, of support. The second important source of support was teaching or research assistantships with 25 to 50 percent of respondents in the several samples listing such assistantships as a major source of funds. However, with respect to assistantships, there was a significant difference between the men and women in our matched sample, with the men much more frequently receiving such aid. For the predissertation stage, the difference was significant at the 6 percent level; for the dissertation stage, at the 1 percent level.

---

[7] These ten fields are those used by the *Journal of Economic Literature*. The percentage of *AEA* members in each field was calculated from the 1974 *Membership Directory*.

Loans represented a relatively unimportant source of funding for women, a significantly more important, though minor, source for men. During the predissertation stage, personal savings represented a minor source of funds for about one-fourth of the women but for almost one-half of the men. At the predissertation stage, support from parents tended to be relatively unimportant in the several samples, except as a minor source for about 20–25 percent of the respondents. No information was solicited about parental support at the dissertation stage. During the predissertation years, employment represented a major source of funds for only 8–10 percent of our respondents. But during the dissertation years, 32–47 percent of respondents in the several samples said employment was a major source of funds.

Support from spouse, as a source of funds, is a relatively complicated variable to analyze. In the predissertation stage, spouse's income was a major source of funds for 15–22 percent of the women and 23 percent of the men; a minor source for 5–10 percent of the women but for 32 percent of the men. For the matched sample, there was a significant difference (at the 1 percent level) between men and women with respect to the importance of spouse's income as a source of support at the predissertation stage. In the dissertation years, spouse's income became more important as a major source for men and women, and more important as a minor source for women, but less important as a minor source for men. But this time there was no significant difference. Part of the reason for the greater incidence of spouse support among men is that a greater proportion of men are married during graduate school. If we look at spouse support for only those who were married during graduate school, we find that at the predissertation stage men still used spouse support more frequently as a source of funds ($p=.01$). However, at the dissertation stage, spouse support was a more important source of funds for women than for men, although the difference was not statistically significant.

Why do women interrupt their Ph.D. studies? About half of the women in the Ph.D. sample interrupted their graduate studies at least once. Of those who indicated how long their interruptions delayed the completion of their dissertation, 17 percent said less than one year, about half (47 percent) said one to three years, 18 percent three to five years and 17 percent more than five years. About 43 percent of the respondents said that having to support their family was a major reason for interrupting their studies; 24 percent said they stopped graduate work for other family-related reasons; and 17 percent said they stopped because of lack of motivation.

Despite interruptions, women Ph.D.s completed their degrees relatively quickly, the mean length of time being 5.7 years. This figure compares quite favorably with the 9.6-year average for the social sciences (male and female 1957–64) cited by Folger, Astin and Bayer (p. 192) and is not substantially different from the 5.3-year average for the men in our matched sample. The 5.7-year figure is even more interesting given the fact that 52 percent of the total female sample were married before completing their degree and one-third had their first child before completing their degree.

Related to the issue of study interruption is the issue of Ph.D. attrition. Why do women Ph.D.s "drop-out"? One hundred and twenty-nine women in our total sample began Ph.D. course work, but did not finish and do not expect a Ph.D. degree.

Because the training of Ph.D.s is extremely costly, from the point of view of institutions as well as individuals, it is useful to inquire into the reasons for Ph.D. attrition. Twenty-seven percent of our "drop-outs" said they left their studies mainly because their interests changed; 23 percent mainly because they decided that, in terms of future financial rewards and/or job opportunities, the work required for the degree "wasn't worth it." Eighteen percent left mainly because of family-related reasons, 8 percent mainly because of an inability to do the required work, and the rest for miscellaneous reasons. When asked "Was there anything that could have been done that would have changed your decision to stop your studies?", 71 percent answered "yes." Financial aid was often (20 percent) mentioned as something which could have changed their decision. But the modal response (36 percent) concerned nonfinancial department support (encouragement, allowing part-time attendance, discussion of job opportunities for Ph.D.s, etc.). Joseph Mooney, in his study of Wilson Fellows' Ph.D. attrition, noted an inverse relationship between graduate school size and the proportion of graduate students acquiring a Ph.D. To the extent that large schools are impersonal and fail to provide much encouragement to students, our finding is related to Mooney's. The discussion of Ph.D. attrition by Michelle Patterson and Lucy Sells also points to the importance of encouragement in decreasing Ph.D. attrition.

III. Conclusion

Given the preliminary nature of our analyses and results, the policy implications of our findings need to be stated cautiously. Yet, for the purpose of reducing occupational segregation, three suggestions can be made even at this early stage of our study.

(1) The high-school years are important for women's commitment to a career. We should arrange more frequent discussions between people in currently male-dominated professions and high-school girls.

(2) Most women become economists because they find that they like the subject matter. Methods of interesting more women in taking some economics courses should be widely discussed.

(3) At the Ph.D. level, women interrupt their studies most frequently in order to help support their families. When they drop out of Ph.D. programs, they feel that better financial aid and more encouragement (financial aid is itself one form of encouragement) could have induced them to continue their studies. The possibilities of significantly increasing teaching and research assistantships as well as departmental support for women graduate students should be seriously considered.

REFERENCES

H. S. Astin, *The Woman Doctorate in America,* New York 1969.

C. S. Bell, "Report of the Committee on the Status of Women in the Economics Profession," *Amer. Econ. Rev. Proc.* May 1973, *52,* 508–511.

———, "Report of the Committee on the Status of Women in the Economics Profession," *Amer. Econ. Rev. Proc.,* May 1974, *53,* 519–523.

A. M. Cartter, *An Assessment of Quality in Graduate Education,* Washington 1966.

J. K. Folger, H. S. Astin, and A. E. Bayer, *Human Resources and Higher Education,* New York 1970.

L. R. Harmon, "Profile of Ph.D.s in the Sciences," *Career Patterns Report No. 1,* Nat. Acad. of Sciences, Nat. Res. Council, Washington 1965.

E. Linn, "Women Dentists: Some Circumstances About Their Choice of a Career," *J. Can. Dent. Assn. 38,* 364–369.

J. D. Mooney, "Attrition Among Ph.D. Candidates: An Analysis of a Cohort of Recent

Woodrow Wilson Fellows," *J. Hum. Res. 3,* 47–62.

M. Patterson, "Sex and Specialization in Academe and the Professions," in A. S. Rossi and A. Calderwood, eds., *Academic Women on the Move,* New York 1973, 313–331.

——— and L. Sells, "Women Dropouts From Higher Education," in A. S. Rossi and A. Calderwood, eds., *Academic Women on the Move,* New York 1973, 79–91.

R. L. Stafford, "A Study of the Process of Vocational Development in Professional Women," *J. of Nat. Assn. of Women Deans and Counselors, 30,* 190–192.

E. Tidball, "Perspective on Women and Affirmative Action," *Educ. Rec.,* Spring 1973.

American Association of University Professors, *Bull.,* Summer 1973, *59,* 215–247.

American Economic Association, "Minutes of the Annual Meeting 1971, Appendix," *Amer. Econ. Rev. Proc.,* May 1972, *51,* 473–474.

American Economic Association, *Directory of Members,* 1974.

U.S. Bureau of the Census, *Census of Population 1970, Occupation By Industry,* Table 8, Washington 1972.

Two Supply Curves for Economists? Implications of Mobility and Career Attachment of Women

By BARBARA B. REAGAN*

The pool of women with highly specialized training is increasing and the tolerance of such women for accepting occupational segregation and narrow views of their career potentials is diminishing. It is crucial that we understand the differences between labor markets for men and women at professional levels, the resulting differences in career patterns, and how barriers to full career development for women impinge on occupational rewards. Only then can we plan effectively for the generation of young women to come. Particularly important at this time is analysis of the market for fields and occupations that are atypical for women.

Overall societal attitudes toward the proper role of women are evolving, yet on the demand side employers' perceived differences as to the employability of women (whether real or imaginary) result in barriers that vary in intensity by occupation. Economics, a stereotypically male profession, has relatively intense barriers for women on the demand side over and above variations in aggregate demand.[1] The demand-side barriers are different from those associated with stereotypically female occupations.

On the supply side, barriers to full career development for women are likely to be those common to all professional occupations plus the effect of women's perceptions of the intensity of the demand-side barriers for the particular profession (i.e., any lack of support of male colleagues, professional isolation and lack of access to information network, or employers' lack of perception of the women's career potential). One group of supply-side barriers includes presence of children, husband's unfavorable attitudes, guilt feelings of the women related to a high sense of responsibility for monitoring consumption at home, and poor earlier education choices based on limited perception of career possibilities. In addition, the two probably most important are geographic mobility or immobility, related to demands of family, and lack of the on-the-job-training, caused by either gaps in the women's career patterns or diminished opportunities for investment in human capital for women when working.

The purposes of this paper are to use preliminary data from the Committee on the Status of Women in the Economics Profession (*CSWEP*), American Economic Association (*AEA*) 1974–75 Survey of Economists to test a model showing that two supply curves exist for economists, instead of one, with the division related to nonpecuniary returns for women, and then consider whether more than two supply curves should be formulated to allow for whether workers are major earners in the family, equal earners, or secondary earners.[2] The model used as a starting point, with subsequent extensions tentatively proposed, is that recently suggested by Nancy M. Gordon and Thomas E. Morton.

---

* Professor of economics, Southern Methodist University. I want to thank Patricia Kirby and Anna Fay Friedlander for help in computing and programming.

[1] For a qualitative discussion of such barriers, see Kenneth E. Boulding and Barbara B. Reagan.

[2] Information on the sampling and survey procedure and quality of sample may be obtained from the author.

I. Comparison of Male and Female Classmates for Highest Degree in Economics

The typical woman economist started work on her Ph.D. when she was a year older than her typical male classmate.[3] The chances are greater that she worked when a B.A. was the highest degree than for her male classmate (62 percent and 44 percent, respectively), and the chances were slightly higher that she worked in a noneconomist job (22 and 14 percent, respectively). She took about half a year longer to finish her Ph.D. than her male classmates, although the difference is not statistically significant with this size sample. This typical woman economist has *not* had a gap of six months or more in her work history since her Ph.D. She has *not* given up a job she liked and moved because her husband changed jobs, moving to another area. Her present salary is $16,500; her male classmate's salary is currently $18,900, or about 15 percent higher than hers.

It has been suggested that persons with large amounts of predegree experience and who take longer to get their Ph.D.s will have a flatter earnings profile and thus a lower salary for any given number of years of experience.[4] The late start (about one and a half years) of the typical woman Ph.D. economist in our preliminary matched sample probably can only explain a small part of the $2,400 difference in pay seven years later.

Untangling the causes and results from differential labor market activity is far more complicated than could be prepared in the three or so weeks since the preliminary data became available and needs the full sample of male economists. However, a few suggestions are possible looking at the first preliminary averages.

II. Model to be Explored

In contrast to previous studies of blue-collar markets, Richard B. Freeman found that the market for college-trained manpower tends to operate "in accord with economic theory, with an informed group of suppliers responding to economic incentives and salaries determined, albeit with a lag, by supply and demand" (p. 36). However, Freeman's study explicitly lays to one side the questions related to women and considers only men in professional occupations.

Gordon and Morton, on the other hand, continued Kenneth Arrow's tradition by considering alternative imperfections in the hiring market and proposed a model of wage discrimination based on the fact that the women's supply curve facing the firm is steeper than the men's. The more steeply

---

Preliminary data from the *CSWEP, AEA* 1974–75 Survey of Economists used in this article are based on the 710 questionnaires returned by 55 percent of the women economists registered with *CSWEP, AEA*, and the first 264 male economists in matched cohorts. (The education cohort was selected at random from three names supplied by each woman economist of men who were in her class when the woman did her highest level of academic work. A name for the employment cohort was supplied by each woman, using an alphabetic criterion, of the male economist at her institution who holds the same rank or approximate level of job responsibility.) Data collection continues for men selected at random for the education cohort from the *American Economic Review, 1974 Directory of Members.* This article, with the exception of the last section, is based on the 52 percent of the women and 67 percent of the men in the education cohort who had completed their Ph.D.s and had worked as economists at least one year. (The numbers of women and men in this matched group are unequal because more of the men matched at the bachelor's or master's level later went on to receive Ph.D.s.) The final section is based on the women and men in the education and employment cohorts who worked in academic jobs in 1974–75.

[3] Subsequent analysis should disaggregate the data by age cohorts to examine how the recent Ph.D.s differ from the more experienced economists.

[4] See G. E. Johnson and F. P. Stafford. They did not consider the effects of fewer on-the-job training or advancement opportunities for women than for men.

sloping supply curve for women that they postulate is based on: (1) differentially perceived nonpecuniary returns from employers by men and women; and (2) higher costs of job search for women than men. The result is that the equilibrium position of the firm is to hire women until the marginal female wage rate, plus the incremental wage bill necessary to bribe male employees to work with the additional women, equals the male wage rate (and marginal revenue product). This model also assumes that men prefer their coworkers or superiors to be men, which widens the resulting wage differential between women and men.

It had previously been noted that incomplete information in the labor market results in transient monopsony power for the firm. Gordon and Morton suggest that permanent monopsony power results from differentially perceived nonpecuniary returns and search costs. Those workers who perceive that a firm will offer greater nonpecuniary returns will work for that firm for a lower monetary wage, thus generating an upward-sloping supply curve. Two types of nonpecuniary returns are suggested to affect women more strongly than men: (1) to the extent that women who are married feel that their job mobility is restricted to the geographic area in which their husbands work, they get nonpecuniary return from an employer in that area and will accept a lower monetary rate from him; (2) to the extent that women have the ultimate responsibility for children, even with household help, women prefer jobs close to their current home more than do men. Even among divorced parents, women still tend to have custody of the children more than men.[5]

III. Test of Model

A. Mobility

Evidence from the *CSWEP* survey for the preliminary group of economists who were classmates for highest degree and subsequently completed their Ph.D.s in economics supports the model that restricted mobility of married women economists is an important factor affecting supply. The fact that for the most part the market for economists is a national one makes mobility all the more an important factor in the supply of women economists. Two-thirds of the women economists with Ph.D.s in the matched sample were married and currently living with their husbands and half had children.

About 60 percent said that spouse's job mobility or immobility had been a major or a minor problem to their career development, with nearly 25 percent saying that it had been a major problem. Furthermore, 30 percent reported that they had given up a job they liked and moved because their husbands or other relative needed to live in another area. More than 10 percent had moved two or more times. Of those who moved for such reasons, the average length of time to find a new job was four months; and 40 percent felt that the new job was worse than the former job, 33 percent felt it was about the same, and 27 percent thought the new job was better.

Often a young married woman is concerned not only about achieving success but also about the effect of her success on her husband. This may particularly be a problem if both are in academics or both are in the same field. "One way of insuring that the academic husband's status will be higher than his academic wife's is to allow the husband's job opportunities to determine where the family lives" (Patricia Albjerg Graham, p. 167).

Men are far less affected by spouse's job mobility or immobility. Less than 25 percent of the male economists had found spouse's job mobility or immobility to be either a major or a minor problem in their career development, and only 8 percent had found it to be a major problem. Furthermore, only 5 percent of the male economists had given up jobs they liked and moved because their wives or other relatives needed to live in another area.

---

[5] In a study of women doctorates, Helen Astin (p. 150) found that "getting adequate domestic help" along with "employer discrimination" were the two greatest barriers to career development, and the third was "husband's mobility." (The first and third are supply-side factors; the second is on the demand side.)

In spite of the fact that so many more women than men feel constrained in their geographic mobility, when the cohorts of classmates were asked whether they would in the next two years be willing to move to a job more than 100 miles away from their present job provided it carried more responsibility and at least 15 percent more salary, 28 percent of the women said "yes" and another 14 percent said "probably yes." Men when asked this question said "yes" 32 percent of the time and "probably yes" 23 percent. Conversely, 45 percent of the men were unwilling to move in the next two years. Thus, it is incorrect to assume, as some employers do, that all women are immobile and all men are mobile.

On the demand side, in employers' minds the effect of perceived immobility of women is much greater than for men. Over a fourth of the women economists with Ph.D.s reported that in the last three years when they have been interviewed for jobs, questions related to their spouses and/or domestic situations appeared to be important considerations in the minds of the potential employer. Only 3 percent of the men so reported.

We have a long way to go before attitudes of male employers change and they are able to consider people as individuals who will themselves measure the tradeoffs involved. Instead, many male administrators are still feeling that they do not want to be party to weakening the bonds of matrimony for someone else's marriage (i.e., interfering with another male's property rights).

B. Care of Children and Home

The nonpecuniary return to women from jobs close to home is shown indirectly by responses to one other question. So far in our society women have the ultimate responsibility for children and home, and this is seen by a large proportion of women as a barrier to career development. More than a third of the women Ph.D.s said that the fact that they had to devote a large amount of time to administering consumption (care of children, house, etc.) had been a major problem in developing their careers fully. More than 70 percent said it was a major or a minor problem. Of the men, however, only 8 percent found such home responsibilities a major barrier to career development.

The 50 percent of the women Ph.D.s who have children average two children per family. The costs of caring for children increase job search costs for women more than for men, as well as being a basis for women perceiving nonpecuniary return from jobs close to home.

C. Distaste

The additional assumption of the Gordon-Morton model that men dislike working with women as peers or as superiors affects the wage differential in that men's wage rate would then be an increasing function of the proportion of women hired. The reasonableness of this assumption for men and women economists is shown by the fact that 29 percent of the women Ph.D.s reported that in the last three years an employer indicated in the female respondent's presence a preference for hiring a man economist rather than a woman economist. In addition, 16 percent of the men surveyed also reported that an employer had stated in their presence a preference for a male economist over a female.

Indirectly, the extent of distaste can be seen from the high proportion of women Ph.D.s who reported that social or professional isolation had been a barrier (56 percent for women and 28 percent for men), and from the fact that 42 percent of the women reported that lack of support from colleagues had been a major or minor barrier to career development. (Only 1 percent of their male classmates who got their Ph.D.s reported that lack of colleague support was a major problem and 14 percent that it was a minor problem).

On the demand side, employers often reflect their understanding of male worker distaste for women colleagues or supervisors by failing to see women as having the potential for special assignments, on-job training opportunities, or professional development into department or agency administrators. It should be noted, however, that some or even much of their reflected distaste may be an erroneous reading of the situation or an oversimplified projection of the administrator's own biases. Nevertheless, about 10 percent of the women Ph.D. economists compared with only 1 percent of their male cohort reported that employer's lack of perception of their potential had been a major barrier to their career development, and another 33 percent of the women, but only 15 percent of the men, reported that it had been a minor barrier.

IV. Extension of Model

A. Discontinuities in Career Patterns

There is another set of circumstances that might work in the opposite direction to those suggested so far in this paper and result in decreasing the slope of the supply curve for women more than for men. If so, this would reduce the wage differential between men and women. This set of factors is related to years of experience and discontinuity in career pattern. To the extent that women at lower wage levels (primarily non-Ph.D.s in academic jobs) have career patterns with major interruptions for childbearing and rearing, the supply of women economists at lower wages is decreased. At upper wage levels the supply might not be decreased because higher salaries permit hiring competent housekeepers and child care arrangements. The net effect, therefore, could be to lower the slope of the supply curves for women. Preliminary data from the *CSWEP* survey suggest, however, that the slope of the supply curve for women economists is not greatly affected by major gaps in their careers.[6] Thus, this factor would not explain much of the male-female wage differential.

Of the economists who now have Ph.D.s, 24 percent of the women had one or more gaps of six months or more in their work history when their highest degrees were B.A.s or M.A.s for reasons other than further education (compared with 2 percent of the men). About five-sixths of these gaps were for demands of home or spouse or personal reasons. The other major reason was "inability to find suitable job." Some of these gaps are undoubtedly tied to moves to new areas when husband moved.

Another 20 percent of the women (but only 1 percent of the men) had gaps of six months or more after they received their Ph.D.s when opportunity costs were higher and higher incomes made paid child care more feasible. The average length of gap after the Ph.D. was two years for the women who had career gaps. Again, most of the reasons were to meet demands of home or spouse and the remainder, no suitable job.[7]

---

[6] Jane W. Loeb and Marianne A. Ferber (p. 247) in their study of the status of women faculty at the Urbana-Champaign campus of the University of Illinois found little difference between men and women with respect to years of professional experience or years of academic experience once age and year of degree are considered. (The faculty study was not limited to those with Ph.D.s.) For women doctorates, a strong pattern of discontinuity was not observed by Astin (pp. 156, 157). Of those women receiving Ph.D.s in 1957 and 1958, 91 percent were in the labor force seven or eight years later. Of these, 81 percent were working full-time and only 10 percent part-time. Nearly 80 percent had never interrupted their careers, and for the rest the median interruption was only fourteen months.

The career commitment of the women Ph.D. economists appears to be as strong as their male classmates who get Ph.D.s. Similar proportions of females and males have worked as economists half or more of the elapsed time since their B.A. degrees (84 and 82 percent, respectively). Over 90 percent of the females are currently members of the *AEA*, compared with 85 percent of the males. Ninety-one percent of the women see their work history as a development of a career rather than just a series of jobs. The corresponding proportion for males is only a little higher, 95 percent; this is not a statistically significant difference.

B. Relative Earning Position Within Family

When sex has been used in past empirical work on labor force participation, it often has been used as a proxy for the relationship of earners within the family with the expectation that women are secondary sources of family income. This is not always the case.

Another fruitful way of looking at labor force participation is to consider the relative proportion of family earnings contributed by the wife and the husband. It might be expected that women who were the major family earners (single or married) would have labor force behavior more like men who are also the major earners (single or married). It might also be expected that the behavior of these major workers (male or female) would differ from that of women who were the minor earners in the family. Furthermore, the behavior of men and women who were essentially equal earners (say, contributing 40 to 60 percent of the family income) would again differ.

Although more refined analysis is needed with the larger sample, a few points are suggested from the first estimates now available. (See Table 1.) These data refer to economists with academic jobs in 1974–75, excluding those with business, government, or other nonprofit positions, but they are not limited to economists with Ph.D.s.

The two groups of women who are major or equal earners are for the most part similar. They in turn differ from the women who are minor earners in the family.[8] The women who are minor earners took longer to obtain their Ph.D.s, nearly always have found the need to devote large amounts of time to administering consumption to be a problem in career development, are not as mobile in terms of the current job market, have more often in the past given up a job they liked and moved because a spouse or relative needed to live in another area, and more often have had gaps in their work history of six months or more. The question immediately arises as to which is the direction of causation. Are their earnings lower because of their slow start, their immobility, their lower on-the-job training, their conflict with home and job demands, or is their labor market behavior in subordinating their career to their spouse's rational behavior considering relative earnings?

The differences in these selected items

---

[7] Gaps in careers affect not only the supply curve but also decrease the amount of on-job training. A high turnover rate in jobs would also decrease the amount and quality of on-job training possible. In the years since they received their Ph.D.s, the women averaged 2.2 jobs, whereas the men averaged 1.4 jobs. However, this difference is not statistically significant at the 95 percent level with this size sample.

[8] Several studies have shown that single women have a less difficult time in career development, i.e., they earn more and are hired for more lucrative jobs.

TABLE 1—LABOR MARKET ACTIVITY AND CAREER BARRIERS BY TYPE OF EARNER, BASED ON RESPONDENT'S CONTRIBUTION TO FAMILY INCOME IN 1973–74

(Economists employed in academic positions in fall 1974)

Item	Major earners (>60%)		Women Minor Earners (<40%)	Women Equal Earners (40–60%)
	Women	Men		
Number in preliminary sample	66	196	27	37
Percent married and currently living with spouse	35	88	100	100
Respondent's proportion of total family income:				
Under 30 percent	0	0	57	0
30–39	0	0	43	0
40–49	0	0	0	54
50–59	0	0	0	46
60–69	58	35	0	0
70–99	25	33	0	0
100	17	52	0	0
Highest degree in economics (percent):				
B.A.	3	4	0	0
M.A.	8	5	7	3
A.B.D.	20	9	7	11
Ph.D.	68	82	85	86
For those with Ph.D.:				
Average number of years to complete Ph.D.	5.0	5.3	7.5	4.8
Average age when completed Ph.D.	30.8	31.2	33.4	29.6
Percent who perceived the following as problems in developing careers fully:				
Need to devote large amounts of time to administer consumption (child care, house, etc.)	56	46	93	72
Major problem	24	8	67	23
Minor problem	32	38	26	49
Spouse's job mobility or immobility	52	22	52	64
Major problem	36	4	26	22
Minor problem	16	18	26	42
Lack of support from colleagues	44	20	20	47
Major problem	13	2	0	12
Minor problem	31	18	20	35
Social or professional isolation on job	57	33	40	64
Major problem	21	7	0	28
Minor problem	36	26	40	36
Current job mobility:				
Percent who would be interested in change of job now	49	44	27	48
Percent who in next two years would be willing to move to a job 100 or more miles from present job providing it carried more responsibility and at least 15 percent more pay than present job	35	35	12	26
Previous job mobility:				
Percent who gave up job liked and moved because spouse or relative needed to live in another area	22	4	46	31
Percent with one or more gaps in work history of six months or more for reasons other than further education:				
With B.A. or M.A. as highest degree	12	6	30	31
After Ph.D.	9	0	26	12
Average length of gap after Ph.D. for those having gap (years)	1.9	—	3.2	4.8

Source: Preliminary data from *CSWEP, AEA* 1974–75 Survey of Economists.

between women and men who are all major earners, either single or married, suggest that sex is still needed as a stratification variable in addition to type of earner. The major career barriers reported by women who are major earners are significantly different from those reported by male major earners. Furthermore, many women economists who are either major or equal earners at this time appear to have subordinated their careers to those of their husbands at one or more previous decision points. Such is the effect of long-standing custom.

V. Conclusions

Preliminary data from the *CSWEP, AEA* 1974–75 Survey of Economists suggest that the supply curve for women economists is more steep than the curve for men because of nonpecuniary returns from employers perceived by women. In addition, men's attitudes toward women colleagues are seen as barriers that have impeded full career development for women economists and thus exacerbated the resulting wage differential. These effects undoubtedly are further deepened by women's reduced will to bargain due to their perception of the effects of previous societal conditioning of colleagues, employers, and themselves. Among women and men economists with Ph.D.s, there are differences in continuity of career patterns both before and after the doctorate is received. These probably are not great enough to change the above conclusion that differences in the supply curve will lead to lower salaries for women, although more detailed analysis is needed.

Supply curve specification should consider not only stratification by sex but also whether workers are (1) major or equal earners in the family or (2) minor contributors to family income.

While more careful analysis of differences in the rewards (salary, rank, and tenure) as functions of various factors must await the larger sample of the male cohort of economists, the tentative conclusions of this analysis are borne out by the reported perceptions of the group studied here: 28 percent of the women economists with Ph.D.s felt that they were paid less or had lower level jobs than they would have had if they had been of the opposite sex.

REFERENCES

H. S. Astin, "Career Profiles of Women Doctorates," in A. S. Rossi and A. Caldenwood, eds., *Academic Women on the Move,* New York 1974, 139–62.

K. E. Boulding and B. B. Reagan, "Combatting Role Prejudice and Sex Discrimination," *Amer. Econ. Rev.,* Dec. 1973, *63,* 1049–61.

R. Fels, ed., *Amer. Econ. Rev., 1974 Directory of Members,* Oct. 1974, *64.*

R. B. Freeman, *The Market for College Trained Manpower,* Cambridge 1974.

N. M. Gordon and T. E. Morton, "A Low Mobility Model of Wage Discrimination with Special Reference to Sex Differentials," *J. Econ. Theory,* Mar. 1974, *7,* 241–53.

P. A. Graham, "Status Transitions of Women Students, Faculty, and Administration," in A. S. Rossi and A. Caldenwood, eds., *Academic Women on the Move,* New York 1974, 163–72.

G. E. Johnson and F. P. Stafford, "The Earnings and Promotion of Women Faculty," *Amer. Econ. Rev.,* Dec. 1974, *64,* 888–903.

J. W. Loeb and M. A. Ferber, "Representation, Performances and Status of Women on the Faculty at the Urbana-Champaign Campus of the University of Illinois," in A. S. Rossi and A. Caldenwood, eds., *Academic Women on the Move,* New York 1974, 239–54.

JOINT LUNCHEON ADDRESS

Inflation and the International Monetary Situation

By H. JOHANNES WITTEVEEN*

Today, I thought it might be appropriate for me to say a few words about the nature and causes of the accelerating inflation from which the world has recently been suffering, and in particular about any interrelationships that may exist between inflation and recent developments in the exchange rate system.

A certain degree of price inflation has been pervasive throughout the postwar period. During the 1950's its importance seemed on the whole to be declining, but later the trend was reversed. During the decade of the 1960's, national income deflators in the industrial countries rose, on the average, at an annual rate of 3.4 percent. While the rate of increase fluctuated somewhat, there was a gradual tendency for inflation to gather speed towards the end of the decade. For a year or two this tendency was checked by the recession of 1970 and 1971, but in the following years the acceleration was rapid indeed with price increases of over 7 percent in 1972 and almost 12 percent in 1973 and 1974.

Developing countries have always, on the average, had higher rates of inflation than industrial countries, but the same trend is visible among them: their consumer price indices, which had risen at an average rate of 13 percent during the 1965–72 period, accelerated to 24 percent in 1973 and 35 percent in the early months of this year.

During the 1950's and 1960's those who deplored what was then called inflation could console themselves with the thought that at least it was accompanied by an unprecedentedly rapid expansion in production and rise in living standards. If the inflation had to any extent served to promote the exceptional real growth, it might seem to some not too high a price to pay. In the 1970's, however, a very sinister phenomenon, previously familiar to us only from the history of hyperinflations, has reappeared and recently become acute, namely, the concurrence of steeply rising prices with declining output and rising unemployment. This phenomenon, sometimes termed "stagflation," presents those responsible for economic policies with an excruciating dilemma.

Looking back, it would seem that the present inflation had begun as the normal expression of an excess demand situation evoked by policy responses to a preceding recession. The source of many of our current problems may be traced to the slowdown in economic activity which occurred in 1970–71 in most industrialized countries. Real output expanded rather sluggishly during these two years, averaging 2.7 and 3.7 percent in real terms for the industrialized countries as a group. Government policy subsequently became more expansionary. With hindsight, it is possible to say that the fiscal and monetary policy actions taken were probably too stimulative. For instance, the average rate of monetary expansion in industrialized countries increased from a relatively moderate

---

* International Monetary Fund.

rate of 7 percent during 1969–70 to over 12 percent in 1971 and 1972. This high rate of monetary growth was accompanied by, and to some extent reflected, the pursuit of expansionary fiscal policies in many countries. To this expansion in the monetary aggregates in the various national economies must be added the sharp increase in credit available through the Eurocurrency market. The estimated size of the Eurocurrency market increased from $44 billion at the end of 1969 to $132 billion, or almost 15 percent of the world monetary stock, at the end of 1973. This must have contributed in no small measure to worldwide inflationary forces.

As a consequence of the stimulative actions taken during 1971–72, the world economy entered a phase of strong expansion in 1972–73. During these two boom years the rate of real *GNP* growth averaged 6 percent, a rate that was not sustainable on a long-term basis. As bottlenecks and capacity constraints were reached, excess demand pressures built up quickly and contributed to the price increases to which we have referred.

Though the current inflation began and gathered strength under the stimulus of demand pressure, there seems little doubt that in its latter stages it has been aggravated and accelerated by cost-push factors, some of which, in turn, have arisen from particular features of the demand inflation. It is these cost-push factors which, in turn, have been responsible for the coincidence of inflation and declining output.

Thus, the boom of 1972 and 1973 had a particularly strong effect in raising prices of foodstuffs and raw materials. Long-run factors, such as rising population and rising per capita income, may have been tending to expand demand for agricultural products faster than the trend growth of supply. However that may be, supplies have recently been curtailed by a series of bad harvests. Consequently, agricultural stockpiles were reduced. Their disappearance removed a factor which had served to mitigate the price-raising effects of past periods of industrial expansion. The elimination of the stockpiles served as a signal to speculators that future shortages would find expression in price changes rather than in sales from existing stocks.

The rise in the price of foodstuffs and most other primary products, though of unusual magnitude, must itself be regarded as an aspect of demand-induced inflation. The same cannot be said of the sharp increase in crude oil prices at the end of 1973 which added another powerful impetus to world inflation. Petroleum is the single most important commodity in world trade and its widespread use as an industrial input, as well as a consumer good, affects wholesale and consumer price indices directly. It has been estimated that the direct impact of the crude oil price increase added some 2 to 3 percent to world prices.

The soaring flight of primary product prices arising from these causes would appear to have stimulated an increase in the prices of finished products greater than could be accounted for by its direct effect on costs. It has done so by evoking increases in other cost elements to a greater extent than could be justified by demand conditions, thus magnifying and accelerating the inflation.

Thus, labor has generally been unwilling to accept the drastic reduction in real income growth—or even absolute decline in real income—which the rise in primary product prices would tend, under market principles, to require of it. And this reluctance has understandably been enhanced by the increases in paper profits which the inflation itself, in conjunction with conventional principles of cost accounting based on historic costs, has tended to bring about.

Another important element in this connection is indexation. Many wages and social benefit payments had been indexed during the late 1960's and early 1970's when we experienced an inflation that was relatively mild compared to our present situation. Often, these contracts contained threshold clauses, providing for increased wages or benefits only after a certain price index had increased by a stipulated percentage. The accelerating price inflation in 1973 triggered many of these clauses and consequently costs to business enterprises and governmental expenditures for social benefit payments increased. At a late stage in inflation, when price expectations, enhanced by an uncertainty factor, are tending to exceed actual inflation, indexation may play a part, in conjunction with demand management, in bringing the process to an end. At an earlier stage, however, its influence almost certainly runs the other way.

As regards profit margins, two opposing forces have probably been at work. Insofar as firms have based their pricing policies on historic cost or insofar as price controls, where they exist, have been administered on this basis, price increases may have been temporarily mitigated or held back. There has, however, been a very natural, and indeed desirable, tendency for firms to shift their pricing policies to a replacement cost basis even where national authorities have been reluctant to reckon taxable profits in this way. On the other hand, the widely practiced system of percentage markups as such is obviously one which tends to multiply the price-raising effects of any increases in costs of materials or labor, whether these increases are a reflection of demand-pull or cost-push factors. Where the total amount available for distribution between capital and labor in industry has been diminished by the rise in primary product prices, the combination of linking wages to the cost of living and the application of percentage price margins by employers sets up a perfect mechanism for the production of a wage price spiral.

The system of direct taxation in industrial countries, both on account of its progressivity and of its adherence in some countries to historic cost accounting principles, tends, of course, to act as a drag on demand inflation. There have been indications, however, that it may be tending to promote cost-push inflation insofar as unions in their wage bargaining and employers in their pricing practices try to maintain a given net-of-tax income. Somewhat ironically, with a progressive tax system the effect of such further inflation will be still further to reduce real after-tax incomes.

As inflation has persisted at an accelerating pace, people have become increasingly more conscious of it, increasingly inclined to expect its continuance, and increasingly desirous of protecting themselves against it. This may even go to the point where economic units not only try to recuperate past losses, but also make an attempt to protect their position against the expected future erosion of their economic position due to inflation by anticipatory price or wage increases. This has naturally given a further impetus to inflation and, together with the factors already mentioned, has contributed to a shortening of the time lags between the various stages of inflation. Thus, we have shorter union contracts, indexation of wage and social benefit payments, flexible interest rate clauses, changes in accounting practices, and other arrangements under which price increases are passed on more rapidly than would otherwise be possible and perhaps even in a magnified form.

All these factors, taken in conjunction, not only contribute to the inflationary spiral but, by accelerating the interaction of price on price, make the system increasingly sensitive to exogenous inflationary shocks. In this context it is easier to understand the magnified repercussions of the increase in oil prices which I referred to earlier.

I do not wish to embark today on the theme of what needs to be done to counter inflationary pressures, or to remedy the stagflation from which we now suffer. I would only sound one warning note, derived from a good deal of International Monetary Fund experience. Partial measures, designed to hold down the price of particular goods and services, such as those which enter prominently into cost of living indices or those produced by the public sector, whether they operate by way of price control or of subsidy, are unlikely to be successful in reducing inflation as a whole and may often stimulate demand. In addition, they usually create scarcities or divert resources from their most appropriate uses. Only comprehensive measures, whether in the field of demand management or of incomes policy or both, will serve.

Thus far, the causes of inflation which I have analyzed have been those which could operate in a closed economy. I should now like to focus on an aspect of the arrangements governing the transmittal of inflationary or deflationary impulses between countries.

It has not escaped notice that the period—from 1973 on—in which inflation has begun to get out of hand in the world is also the period during which floating exchange rates have been the rule rather than the exception among industrial countries. The question is whether this is a chance combination of circumstances or whether there is an interrelationship between the two phenomena and, if so, what it is.

First, a few words on the role of inflation in the abandonment of the par value system.

As we have seen, a moderate degree of inflation prevailed throughout the postwar period, becoming less moderate in the late 1960's. Moreover, the character of the inflation changed somewhat, cost-push elements becoming more prominent. Whereas earlier inflation had been fueled by reconstruction needs and sometimes by the incapacity of governments to finance their expenditures without recourse to the banking system, later the prime motive was to preserve a high level of employment which, given more aggressive behavior of labor, could only be done at rising price levels.

Stable exchange rates can only be combined with desired levels of employment when there is a consistent development of price levels in different countries. Now, abstractly, the mere fact that prices are rising everywhere does not prevent relative national price levels from remaining constant or moving in a direction consistent with international equilibrium. In practice, however, the higher the degree of inflation, the larger are likely to be the divergences in inflation rates. Moreover, there is no reason why such divergences should occur in the direction required by structural changes affecting the balance of payments.

The growing importance of the cost-push element in inflation probably tended to accentuate these divergences and to necessitate exchange rate adjustments. While profits were high and unions unaggressive, countries in payments difficulty could get back into balance by slowing down the expansion of demand without undue effects on employment; when cost-push became the driving element in inflation, this ceased to be true and devaluation became the only solution.

Of course, exchange rate adjustment could take place under the par value system. During the 1960's such adjustments were often unduly delayed, but later, when they became more frequent, the volume of anticipatory capital movements grew too great to be contained within the structure of this system, and resort to generalized floating became inevitable.

A question which is perhaps more controversial than that of the role of inflation in compelling a transition to flexible exchange rates is that of the possible role of the flexible exchange rate system in promoting inflation.

There are various reasons, much discussed in the literature, for thinking that exchange flexibility and, in particular, freedom from the obligation to adhere to any particular exchange rate should make it easier for a country to cope with inflationary pressures and to preserve internal financial stability.

In the first place the option of allowing one's currency to appreciate should enable the monetary authorities of a country to buffer it against the impact of demand pressures from abroad, whether these arise from lax demand policies practiced in other countries or from structural shifts in foreign trade, though in the latter case adjustments will still be required as among particular industries. It should also enable them, though at some cost in terms of exchange stability, to protect the monetary system against an influx of reserves that would otherwise be difficult to offset. The degree of insulation against external shocks that is achieved in this way is reduced by a number of factors. Thus, the more inflationary countries may for a time attract a sufficient volume of investable funds to prevent their exchange rates from falling, and those of the less inflationary countries from rising. Again, the low level of foreign trade elasticities in the short run means that when exchange rates do move the buffer against external demand variations will be fully effective only after a time lag. Nevertheless, the monetary insulation will operate immediately, and I think that countries like Germany or the Netherlands would feel that the regime of floating rates had served to protect them, at least in some measure, against imported inflation. At the same time, it has to be borne in mind that insofar as this mechanism tends to prevent the spread of inflation it also tends to concentrate and magnify it in the countries in which it originates.

A second, though related, advantage of the flexible rate system from the standpoint of stabilization policy is that the instruments of demand management, particularly monetary policy, operate much more powerfully on the domestic economy when they are not subject to the leakages that affect an open economy under fixed exchange rates. Monetary contraction need no longer be frustrated by an influx of reserves and money from abroad and will, by promoting exchange appreciation, exercise a deflationary effect on the current account balance of payments. Here again, some of the deflationary effect will be achieved at the expense of other countries, but these countries have both the incentive and the possibility to follow suit with effective stabilization policies of their own.

Balancing these arguments in favor of flexible rates, there are two arguments frequently advanced against flexible rates from the standpoint of stabilization policy. There is, first, the contention that governments, especially democratic ones, are naturally prone to pursue expansionary policies and that exchange flexibility removes the inhibitions formerly imposed upon this tendency by fear of running out of reserves and of incurring the political odium of devaluation. There is some force in this contention, though the fact that under floating rates the effects of overexpansion must be kept at home with immediate repercussions on the domestic cost of living should provide, to some extent, a substitute discipline, if possibly a less effective one. I must add that there is not much evidence that the transition to floating rates in 1973 has led to a perceptible loosening in demand management. The expansion in money stocks which laid the foundation for the boom took place from 1970 to 1972 which, despite an interlude of floating in 1971, was on the whole a period of fixed rates.

Another familiar argument against floating rates which seems somewhat dubious in practice is the contention that spontaneous inflation pressures within a country, which would, under fixed rates, have been dispersed harmlessly over a wide area where they could be more easily dealt with, might, under floating rates, be "boxed in," as it were, in the country of origin in a way that makes them more difficult to control. This might have some validity under a system of free floating when countries were forbidden to intervene to support the exchange rate, but under the system of managed floating which actually prevails, countries are free to intervene in support of the rate and when domestic inflationary pressures are strong are very likely in practice to do so.

There are, however, some other reasons for suspecting that floating rates may on balance have contributed to the average rate of inflation among countries in the last year or two. Most of these operate from the side of costs rather than of demand. For example, there is the question of the risks attached to international transactions and to the production of internationally traded goods. We have seen some rather large fluctuations in exchange rates since generalized floating began, particularly as between the *U.S.* dollar on the one hand and the currencies participating in the *EEC* snake on the other, where the amount of stabilizing official intervention has been rather small. Thus, the deutsche mark appreciated by some 30 percent vis-à-vis the dollar from February 14, 1973 to July 1973, then depreciated by 22 percent by a date in January 1974. Five months later, by May of 1974, the mark had risen again by 20 percent to a new peak only to fall 10 percent to a new trough in September. Since then it has risen again by nearly 10 percent. It can be argued that these fluctuations have shown a declining tendency. Nevertheless, they have been large and must have added to the risks of doing business, not only as between the two countries in question, but among all those countries whose currencies are tied either to the dollar or to the mark, or even those which try to steer a middle course between the two.

While one can hedge in forward markets against some of these risks, these markets are not well developed for all relevant maturity periods, and the spreads charged on forward contracts are relatively high. Besides, production of international trade goods or investment in such production involves many types of exchange risks which cannot be hedged by short-term forward contracts. Of course, devaluation and revaluation risks would exist under a par value system also; moreover, in some cases the adjustment of exchange rates facilitated by a flexible system will serve to offset other risks, such as the risk of divergent developments in domestic currency costs. But the sort of short-to-medium term fluctuations in rates to which I have drawn attention have characterized the flexible rate system up to now, and it would be strange if required profit margins had not been widened to cover such risks.

These fluctuations probably promote inflation in yet another way. Exchange rates influence the distribution of demand between domestic and foreign products, and rate fluctuations will involve demand shifts which will later in some measure be reversed. These demand shifts may well exercise a ratchet effect on inflation. Given the downward rigidities in wages and prices characteristic of modern industrialized economies, shifts in purchase patterns brought about by the exchange rate changes, or even the expectation of such shifts, may tend to raise prices in the countries of depreciating currency without effecting a corresponding price reduction in the countries of appreciating currency. It might be argued, in view of the slowness with which demand patterns respond to exchange rate changes, that this ratchet effect would not be very important for fluctuations as short term as those to which I have just referred. However, there is evidence that prices are adjusted to meet changes in competitors' prices evoked by exchange rate movements in advance of the expected shifts in demand, and that these anticipatory price adjustments are more marked in the upward direction.

This brief review of the interrelationship of inflation and exchange rate flexibility seems to yield some tentative conclusions. In the first place, the influence of inflation on the exchange rate regime has been more considerable than the influences operating in the opposite direction. As regards the bearing of the exchange regime on inflation, I would say that floating makes it easier for countries with relatively strong anti-inflation policies to achieve the desired results, but that the price for this is more inflation in countries already inflating more strongly. Thus, floating which is partly caused by differences in rates of inflation in its turn increases these differentials.

With respect to the average rate of inflation in all countries taken together, floating exchange rates may have been a slight additional fillip to inflation, particularly through their effect on the risk element.

An important distinction has to be drawn here between the role of the floating technique in permitting prompt adjustment to underlying trends or even to cyclical reactions in the balance of payments and its role in permitting shorter term fluctuations. The first aspect is almost entirely beneficial from the standpoint of inflation control, as indeed from the aspect of adjustment in general. The second aspect is the one that may cause trouble both for stabilization policy and, indeed, for adjustment also.

In a decision adopted last June the Executive Directors of the International Monetary Fund have agreed on a set of guidelines or principles which its members will attempt to follow to avoid disruptive exchange rate movements under floating. The course of action which the guidelines encourage is that of smoothing out day-to-day and week-to-week fluctuations in rates, resisting overrapid movements of rates in any particular direction, and resisting undue deviations in rates from their medium-term norm, if such can be discerned. This would tend to eliminate the additional risk resulting from excessive short-run fluctuation in exchange rates.

In the further development and application of these guidelines I therefore see a possible remedy for any net tendency that may exist for the floating rate system to promote inflation. This argues strongly for the development of the international consultation and cooperation which could make these guidelines effective.

UNDERSTANDING WORLD INFLATION

Monetarist and Keynesian Models of the Transmission of Inflation

By WILLIAM H. BRANSON*

The international monetarist model has a fully endogenous money supply, with the domestic rate of inflation fixed by "the" world rate. A standard post-Keynesian model, on the other hand, includes as endogenous variables both the rate of inflation and the balance of payments, with the monetary authorities controlling the money stock. In this paper I argue that the Keynesian model is more consistent with the facts, and that both the monetary and monetarist models are extensions of the Keynesian model with additional, unrealistically restrictive assumptions.

I. The Monetarist Approach

The monetarist theory of world inflation, and its connection to the balance of payments, goes roughly as follows. The excess of the growth rate of the world money stock over the growth rate of output determines the world rate of inflation. By an extension of the "law of one price," this determines each country's rate of inflation. Given this rate of inflation, and interest rates and output pegged by small-country assumptions, movements in the domestic credit base determine reserve changes. The keys to all this are the condition that money demand equals supply, and the extreme small-country assumptions.

The model can be set out fairly simply, following H. G. Johnson (ch. 3). At the level of the individual country we have a demand function for money,

$$M^D = P\cdot m(y, r), \tag{1}$$

and a supply equation $M=R+B$, where R is international reserves and B is the domestic credit base. Here for convenience I assume a 100 percent reserve system. P is the price level (GNP deflator), y is real output, and r is the interest rate. Equating supply and demand yields the equilibrium condition

$$R + B = P\cdot m(y, r). \tag{2}$$

The small-country assumption for capital markets pegs r to "the" world interest rate. The small-country assumption for goods markets pegs P and y. The country is a price taker on world markets, pegging P, and it can sell all it chooses at P, pegging y. Now with P, y and r fixed by the small-country assumptions, movements in the domestic credit base B determine movements in reserves R, on a one-for-one basis.

The final step is the link from the growth rate of the world money supply to the world rate of inflation. Using a subscript w to denote world level variables, the monetarist story has a world supply and demand equilibrium,

$$M_w = P_w\cdot f(y_w). \tag{3}$$

Here the interest rate disappears from the demand function for money. Taking the world money supply and world output (!) as exogenous, the rate of inflation is given as

---

* Professor of economics and international affairs, Princeton University.

$$(4)\qquad \left(\frac{\dot{P}}{P}\right)_w = \left(\frac{\dot{M}}{M}\right)_w - \epsilon_y\left(\frac{\dot{y}}{y}\right)_w,$$

where $\dot{X} \equiv dX/dt$, and ϵ_y is the income elasticity of the demand for real balances. With (4) determining the world rate of inflation, P (or $\dot{P}/P$) in (2) is fixed, and changes in the growth rate of B simply change the path of R.

This monetarist model has at least two empirical implications. One is that national inflation rates converge to the world rate. The second is that reserve changes offset one-for-one changes in the domestic credit base, with the money stock being completely endogenous. The latter implication is not consistent with the empirical evidence. For example, P. J. K. Kouri and M. G. Porter obtained the following estimates of the R reaction to a B change (from estimated capital flow equations), as a fraction of the initial B change: Germany, -0.77; the Netherlands, -0.58; Australia, -0.47; Italy, -0.43. In the monetarist story, all of these should be insignificantly different from -1.0. R. J. Herring and R. C. Marston obtained similar results for Germany and Canada, and the evidence is accumulating on this point. The evidence on the first implication of convergence is shaky, but here another problem arises: a Keynesian-Phillips view of the world yields the same implication of convergence, as I show below in section III.

One reason why the monetarist model is not consistent with the data thrown up by our short-run world is that the assumptions are not consistent with reality. Even in small open economies we would see "the" interest rate and price level varying relative to the world interest rate and price level, if we could observe such variables in the first place. And we certainly see variations in real output around its trend or full employment path even in small countries. So to understand how inflation spreads, we must go to a model that relaxes these monetarist assumptions. My favorite candidate here is the bare-bones post-Keynesian model of A. S. Blinder and R. M. Solow, with a foreign sector added. The monetarist model is a special case of an augmented Blinder-Solow model with addition of the extreme small-country assumptions.

II. The Keynesian-Monetary Approach

The bare-bones post-Keynesian model has aggregate demand determined by the "textbook" *IS-LM* framework, an aggregate supply curve, and equilibrium reached at the intersection of demand and supply. The internal equilibrium so determined can be related to a schedule showing r, y pairs that yield balance of payments equilibrium to determine the external situation.

The supply curve is simply $P = P(y, \beta)$, with $\infty \geq P_y > 0$. This assumes wage rigidity or money illusion in the labor supply function. Here β is a shift parameter reflecting changes in availability of raw materials, with $P_\beta < 0$. A useful improvement here would be integration of the "Scandinavian" two-sector model of price determination described by O. Aukrust.

Aggregate demand is given by the intersection, at a given price level, of the *IS* curve

$$(5)\qquad E(y, r, \tau) + g + b(P, y) = y,$$

and the *LM* curve

$$(6)\qquad M = P \cdot m(y, r).$$

Here E is private expenditure, τ is the tax rate, g is real government purchases of goods and services, and b is the current account balance. Given g, τ, M and P, (5) and (6) determine r, y at the *IS-LM* intersection. As P rises, the y so determined falls, giving a negatively sloped aggregate demand curve, as shown by C. M. Gambs. The intersection of the supply and demand schedules determines equilibrium P, y, and the interest rate comes from the *IS-LM* intersection with *LM* positioned by the equilibrium value of P.

To determine the external situation, the *IS-LM* intersection has to be related to the r, y line giving balance of payments equilibrium, the *BP* line of Figure 1. If the r_0, y_0 point lies below *BP*, there is a deficit (r is too low, given y). The workings of this model are rigorously derived in Branson (1974).

Several implicit assumptions differentiate this model from the monetarist model. Here it is assumed that the domestic monetary authorities control M—that sterilization of the external balance is possible. Further, it is assumed that home goods and assets are either (a) imperfect substitutes for foreign goods and assets, or (b) partially nontraded. This gives some freedom of movement to P, y and r.

With this framework we can look at the effects of changes in exogenous variables on the home price level and the external balance, which are determined endogenously. An increase in the money stock shifts the *LM* curve, and the demand curve in P, y space, out. This pulls the price level up and increases the balance of payments deficit by moving the *IS-LM* intersection away from the *BP* line. From the point of view of other countries, this will appear as normal imported inflation. Their exports and price level will rise, and they will register a payments surplus.

The effects of fiscal policy are not so clear. An increase in government purchases or a tax cut shifts the *IS* curve and demand curve out, increasing y and the price level. But the effect on the balance of payments is unclear, as noted originally by J. M. Fleming. If the *BP* line is steeper than *LM*, the expansionary fiscal policy increases the deficit; if *BP* is flatter than *LM*, a reduction of the deficit results. Thus from the point of view of other countries, the imported inflation due to an expansionary fiscal policy in the originating country could be accompanied by a surplus or a deficit in the balance of payments.

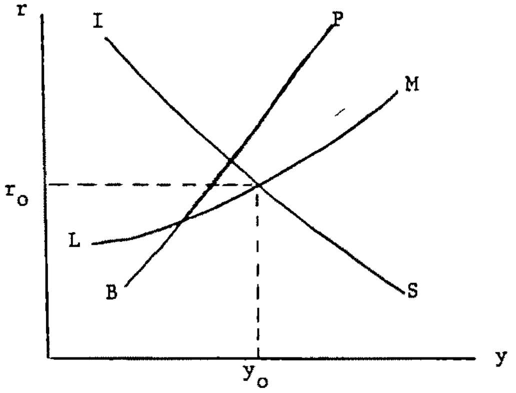


FIGURE 1: INTERNAL AND EXTERNAL BALANCE

The empirical results on the slope of the *BP* line generally tend to favor the case of relative capital immobility: *BP* steeper than *LM*. For example, my calculations (Branson 1975) using the S. Y. Kwack current account model and the Branson-R. D. Hill capital account equations for the United States show a *BP* line significantly steeper than *LM*. J. F. Helliwell reports simulations combining the MIT-Penn-SSRC (*MPS*) model of the United States and the Canadian *RDX2* model. These show a *BP* curve for Canada with about the same slope as *LM*; an expansionary fiscal policy in Canada is neutral with respect to the exchange rate or the balance of payments.

Thus it does appear that inflationary impulses emanating from shocks in the real demand sector in the United States are accompanied by *U.S.* deficits and surpluses in other countries, but only because of the relative immobility of capital in the

U.S. balance of payments. In general, the message of this simplest Keynesian model is that relationships between the balance of payments and the price level are complicated and not one-signed. If we know the source of an exogenous shock, we can probably determine what it will do both to the domestic price level and to the balance of payments. But knowing what it does to the balance of payments does not tell us what it does to the price level; we have to know the model and the source of the shock for that.

The Keynesian model described above is a pure flow-equilibrium model. It takes the stock of money as given, assuming that the authorities can sterilize the effects of the balance of payments on the money stock. That this is generally the case is suggested by the findings of Kouri-Porter and Herring-Marston, cited above. If we relax this assumption, and allow the foreign balance to affect the money stock, we come to the "monetary approach" to the balance of payments.

If an external imbalance is not sterilized, then as long as the *IS-LM* intersection is not on the *BP* line in Figure 1, the money stock will be changing, moving the *LM* curve, and moving the *IS-LM* intersection toward the *BP* line. If the internal balance point is below the *BP* line, there is a payments deficit, the money stock is shrinking, and the *LM* curve is shifting left, tending to reduce the deficit. The system comes to a stock equilibrium when the internal balance point lies on the *BP* line. This monetary approach is thus a simple extension of the bare-bones post-Keynesian model to adjustment of one stock—money—to equilibrium. The monetarist approach of section I is a further simplification, adding the small-country assumptions. It is only in that case that clear results for the transmission of inflation appear.

III. Convergence in a Keynesian-Phillips Model

One implication of the monetarist model is the convergence of national rates of inflation toward the world rate. One can obtain the same result from a Keynesian foreign trade multiplier model plus a price-Phillips curve, as well.

Begin with a price-Phillips curve,

$$\dot{P}/P = h(y), \qquad h_y > 0. \tag{7}$$

Here the rate of inflation is an increasing function of real output, given potential output. Next add a trade balance b = net exports, which is a decreasing function of the ratio of home to world prices:

$$b = b(P/P_w), \qquad b_p < 0. \tag{8}$$

Finally, holding constant real government spending and real assets, the foreign trade multiplier implicit in the Keynesian model can be summarized as

$$y = y(b), \qquad y_b > 0. \tag{9}$$

This completes the dynamic system. The trade balance depends on P, y depends on the trade balance, and y drives $\dot{P}/P$. Combining (7), (8) and (9), we have

$$\dot{P}/P = h(y(b(P/P_w))) = \lambda(P/P_w); \tag{10}$$

$$\lambda_p = \overset{+}{h_y}\cdot\overset{+}{y_b}\cdot\overset{-}{b_p} < 0,$$

as required for stability.

The system is described in the simple phase diagram of Figure 2. There the horizontal axis measures y and b in suitable units, and the rate of inflation $\dot{P}/P$ is on the vertical axis. First we locate the y^*, b^* values that yield $\dot{P}/P = \dot{P}_w/P_w$; these are equilibrium values. Then consider an initial point with y_0, $b_0 < y^*$, b^*. At that point $\dot{P}/P < \dot{P}_w/P_w$, so the trade balance is increasing, pulling up y and $\dot{P}/P$ toward the equilibrium point, as shown in Figure 2. If we began with y, $b > y^*$, b^*, we would see movement back

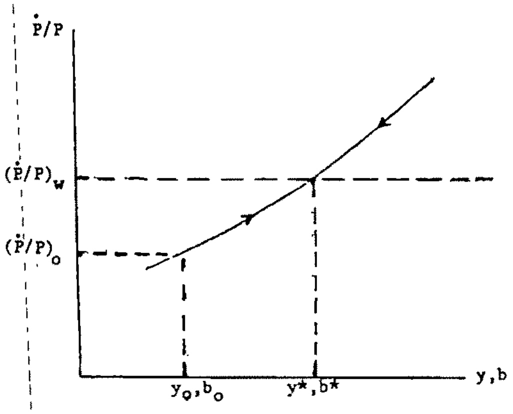


FIGURE 2: HOME RATE OF INFLATION CONVERGES ON THE WORLD RATE

down the phase line. Thus in the absence of policy action, the Keynesian-Phillips model would predict that national inflation rates would tend to converge toward the average world rate, through the action of the foreign trade multiplier.

This line of analysis yields two interesting implications. First, implicit in the Keynesian-Phillips view of inflation is a tendency (a) for rates of inflation to converge, and (b) for real output and unemployment rates to adjust so that rates of inflation converge. This happens through the foreign trade multiplier. Second, since inflation rates tend to converge on the Keynesian-Phillips view, evidence that they are converging, taken alone, cannot be used to support the monetarist view. Thus research designed to distinguish between Keynesian and monetarist adjustment mechanisms must go beyond the study of converging rates of inflation to model the actual mechanism behind such convergence, if it in fact occurs.

REFERENCES

O. Aukrust, "Prim I: A Model of Price and Income Distribution in an Open Economy," *Rev. of Income and Wealth,* Jan. 1970.

A. S. Blinder and R. M. Solow, "Analytical Foundations of Fiscal Policy," in *The Economics of Public Finance,* Washington 1974.

W. H. Branson, "International Transmission of Inflation: A 'Keynesian' Approach," presented at Brookings Institution Conference on Worldwide Inflation, Nov. 21–23, 1974.

———, "Comment" on Kwack and Schink in L. R. Klein and G. Fromm, eds., *The Brookings Model: Perspectives and Recent Developments,* Amsterdam 1975.

——— and R. D. Hill. *Capital Movements in the OECD Area,* Paris 1971.

J. M. Fleming, "Domestic Financial Policies under Fixed and under Floating Exchange Rates," *IMF Staff Papers,* Nov. 1962, *9,* 369–79.

C. M. Gambs, "A Note on Macroeconomic Textbooks: The Use of the Aggregate Demand Curve," *J. Econ. Lit.,* Sept. 1974, *12,* 896–98.

J. F. Helliwell, "Simulation of Canada–U.S. Interdependence: Effects of Stabilization Policies," paper presented at International Seminar on Public Economics, Williamstown, Mass., June 1974.

R. J. Herring and R. C. Marston, "The Monetary Sector in an Open Economy: An Empirical Analysis for Canada and Germany," Rodney White Cent. work. pap., Univ. of Pennsylvania 1974, 7–74.

H. G. Johnson, *Inflation and the Monetarist Controversy,* Amsterdam 1972, ch. 3.

P. J. K. Kouri and M. G. Porter, "International Capital Flows and Portfolio Equilibrium," *J. Pol. Econ.,* May/June 1974, *82,* 443–67.

S. Y. Kwack and G. R. Schink, "A Disaggregated Quarterly Model of U.S. Trade and Capital Flows," in L. R. Klein and G. Fromm, eds., *The Brookings Model: Perspectives and Recent Developments,* Amsterdam 1975.

Understanding World Inflation

By GEORGE L. PERRY*

As the title of this session indicates, we economists are now grappling with the problem called "worldwide" inflation before having really understood "plain old" inflation. Complicating an unsolved problem is seldom a fruitful research strategy. Yet there is no avoiding the global scope of today's inflation and the connections between national economies that may contribute to it. Financial flows among nations complicate domestic monetary policies. Trade transmits the ebb and flow of aggregate demand from one nation to others. The prices of goods traded internationally move together. But such relationships could and have been deflationary as well as inflationary. The crucial questions are whether and why these relationships have been systematically inflationary to each country's price level; and more important, why the problem is inflation in the first place.

Today's inflation is complex, not simple; it has diverse causes, not a single cause; and it is rooted in institutional characteristics of economies that vary in important ways from country to country. It helps to sort some of the major economic developments of recent years into three aspects of the inflation process: its origin, its transmission, and its propagation. These categories are not analytically precise, but they do represent distinguishable parts of the process that has led to today's price performance in different countries.

Price increases have originated in several distinct ways. In the United States, domestic inflation originated in the late 1960's as excess aggregate demand led to overly tight labor markets and more rapid rates of wage and price increase. The food price explosion, which has been a large part of the 1973 and 1974 price increase story, resulted from a global imbalance of supply and demand that came from disappointing weather and wayward anchovies. The formation of an international cartel that quadrupled the world oil price is responsible for another large part of the observed increase in price levels since the fall of 1973. In late 1972, the coincident boom in industrial output around the world started other raw materials prices soaring. These are all very different initial sources of price increase. Yet all generated a part of what we call the inflation of recent years.

The transmission of economic developments across national boundaries takes place through several channels, including export demand, financial flows, and goods prices. There are obvious examples of each channel transmitting inflationary disturbances in recent years. The original price increases just identified directly meant higher prices for importing nations. The decline in the *U.S.* trade surplus, which eventually became a large deficit in the early 1970's, had its counterpart in rising exports from other nations. During this same period, a flood of *U.S.* dollars found its way to European capital markets.

According to the analytic categories I am describing, these transmissions are not the crucial step in the global spread of inflation. Whatever the origins of a price rise, or however inflationary tendencies are transmitted across national boundaries, the process would be limited if inflationary disturbances did not propagate further inflation. But in fact they have, and in a variety of ways, some reflecting structural characteristics of economies and some reflecting policy actions or inactions. In the United States, after having fallen behind average wage changes during the late 1960's, powerful unions caught up in the early 1970's, long after labor demand had eased. In most countries, rising export demand that accompanied the deterioration in the *U.S.* trade position was not offset by revaluation or more restrictive fiscal-monetary measures. Foreign monetary authorities did not neutralize the effect of dollar inflows to European countries and thus permitted them to result in a large expansion of domestic liquidity. Finally, in most countries, the food and fuel price explosions of recent quarters have apparently led to soaring wage increases, which have translated the food and fuel price increases into rising costs and prices for domestically produced output.

---

* The Brookings Institution.

Some would argue that all this describes symptoms of inflation rather than inflation itself, that it points out the places that get flooded while missing the fact that the dam broke. It is convenient to model the world economy by letting crop failures, cartels, and devaluations determine relative prices while the price level and inflation are determined by something simple like the money supply. But nobody should pretend that model answers any of the hard or interesting questions. Even if M_1 could tell us all we need to know about nominal incomes, knowing the growth of nominal income does not tell us how much of it would be real output and how much a rise in the price level. Surely the answer was different after the oil cartel formed than before. It was also different in 1960 from 1970. And what the possibilities are for 1975 is a source of considerable uncertainty, interest, and importance today. What we need to understand is why economies are inflation prone so that restrictive fiscal-monetary measures to combat inflation are so costly and unpopular.

In recent years, with only a few exceptions, countries have used neither fiscal-monetary restraint nor currency revaluation with enough force to hold down prices. It has been argued that European governments were powerless to mop up the excessive liquidity caused by dollar inflows, so they had insufficient control over nominal aggregate demand in their own countries. A refinement of this argument brings in the Eurodollar market and multinational firms as institutional impediments to the conduct of national aggregate demand policy. It has also been argued that currency revaluations are difficult owing to the political strength of export industries. Such assertions may contain some truth and may help explain some policy actions or inactions. But the inflationary bias in supply conditions of most economies and the tradeoff dilemma that this poses for aggregate demand policy are the more fundamental problem. If more fiscal-monetary restraint would have prevented inflation without a politically unacceptable cost in jobs and output, it is likely we would have seen more of it.

Several strands of thought have been developed in recent research on the inflation-unemployment tradeoff, most of them offering explanations for why the tradeoff has worsened. For the United States, it is now fairly well established that demographic shifts in the composition of the labor force mean labor markets are tighter today for any given aggregate unemployment rate. Women and younger workers of both sexes, on the average, enter and leave the work force more frequently than adult males and change jobs more frequently, encountering more spells of unemployment in the course of these transitions than workers with more permanent job attachments. In addition, some researchers have argued that the real cost of unemployment to workers has fallen as a result of more generous welfare and unemployment benefits, thus worsening the tradeoff by reducing the downward pressure that unemployment exerts on the average wage level. On my reading of the evidence for the *U.S.* economy, the first of these effects is noticeable in size, while the second is of uncertain importance. The same effects may have had much different importance in other countries.

Once allowance has been made for factors such as these that may have shifted the inflationary implications of any particular unemployment rate, several additional questions remain to be settled concerning the relation between inflation and labor markets. Much recent research has emphasized the importance of long lags in the inflation process as another explanation of why the short-run tradeoff has worsened. Excessive tightness in labor markets initially has a modest effect in speeding up wages; but with the passage of time, this effect is substantially magnified and wages end up rising much faster than current labor market conditions alone can explain. This kind of effect is often associated with the accelerationist literature, but it is also a part of many more general models of the labor market. It has been advanced both by economists with an atomistic view of the labor market and the way wages are determined and by economists who assign a major role to noncompetitive market power determination of wages. A lot of inflation modeling in the United States has worked with the competitive model, while Scandinavian and British economists have often stressed the importance of market power by labor unions.

Long lags have been hypothesized either through a wage-wage interaction or, more commonly, through a price-wage interaction. Whether a wage-wage or price-wage mechanism is at work is important: recently price indexes have accelerated everywhere for reasons unrelated to wage costs. Presumably we should have expected continued wage moderation on a wage-wage view of the world, but a sharp acceleration of wages on a price-wage view. This also raises another crucial distinction. If we believe a price-wage view, what is the relevant price? Is it the consumer price index which measures the real wage in terms of purchasing power of workers? Or is it a product price index, based on the output produced by labor, which measures the value of labor's marginal product? These distinctions are generally glossed over when lagged effects in the inflation process are simply attributed to expectations. Some sort of distributed lag can be fitted to almost any variable designed to measure wage or price increases and will give a tolerably good fit to the data without addressing any of the crucial structural questions about what actually determines the rate of wage increase.

Many of these unanswered issues about the generation of inflation are sharpened when we focus on the propagation of inflation. First, if the wage-wage model is appropriate, then rising prices of imported goods will not be propagated through faster wage increases into rising prices on domestic production. Alternatively, if the prices of goods labor produces matter in wage determination, then higher import prices could affect wages in the import-competing sector by raising prices in that industry. A further propagation of domestic inflation would then depend on whether a wage-wage mechanism existed that would extend the wage inflation to all of domestically produced output. Finally, if it is living costs that directly affect wages, then higher import prices would directly generate higher domestic wages and, hence, higher prices for domestic production.

Some of these issues can be viewed against the *U.S.* experience of the last couple of years, when sharp divergences among different price measures offer an unusual opportunity to discriminate among different views of the inflation-propagation process. Devaluation did appear to raise the price of Pintos and similar American car makes by raising the price of foreign small cars with which the Pinto competed. And, of course, prices of domestic oil and alternative energy sources rose with world oil prices. How much more of such a pricing response occurred is obscured by the fact that the pricing rules under Phase III of the Wage-Price Control Program were relaxed in 1973 and the fact that some *U.S.* industries, such as steel, were operating at or near capacity that year. Some part of nonfarm price increases during the past two years has come from these sources rather than directly in response to the higher price of competing imports. Nonfarm prices had fallen behind their historic relation to standard unit labor costs during Phases I and II of the wage-price program, and the restoration of precontrol margins, together with the continuing rise in standard unit labor costs, would have predicted about all of the 15 percent rise in average nonfarm, nonoil prices that has occurred in the two years through the third quarter of 1974.

If the main story in the recent *U.S.* price acceleration is food, fuel, and post-control catchup, the main story in *U.S.* wages is relative moderation and very little propagation of inflation through any price-wage-price spiral. Except in a few strong unions that established significant escalator clauses, wages have not responded to the consumer price increases of 1973 and 1974 in anything like the way an elasticity of 1.0 on living costs would have predicted. In fact, real compensation per manhour has actually declined in each of the last six quarters. Unless wage developments in 1975 reverse the verdict, this will be potent evidence against a strong price-wage mechanism working through living costs. Not only has the soaring price of food and fuel not been converted into correspondingly faster rates of wage increase, but even rising prices of nonfarm (nonoil) output have not thus far been translated into a corresponding acceleration of money wages.

The structural response of wages in other countries appears to have been substantially different from this. The striking acceleration of money wages elsewhere suggests either that their labor markets have been much tighter than those in the United States and, indeed, much tighter than in the past experience of most of these countries themselves, or else that the response of wages in some other countries to outside price increases is very different. I suspect the latter explanation is the more important part of the story. Over the most recent twelve- and twenty-four-month periods, it appears that average wages have risen faster than consumer prices in Great Britain, France, Germany, Japan, and Italy, although I must warn that I am uncertain about the reliability of the wage measures behind these calculations. If these measures are close to correct, these countries have a wage-determining structure that is much more closely indexed to consumer prices and more inflation prone than the process in the United States. Understanding the reasons for such differences would be an important step in understanding what makes economies inflation prone.

Governments are now going down the road of choking off aggregate demand to try to stop inflation. On the evidence, fiscal-monetary authorities are confronted with an impossible assignment in trying to maintain both prosperity and price stability in the face of substantial inflationary biases in their economies and inflationary shocks from abroad. Understanding inflation comes down to understanding these biases. And really curing inflation cannot mean simply causing enough unemployment to make it stop, but rather must involve reducing the biases and finding ways to make our institutions less inflation prone.

In the United States, fruitful research was being done in recent years aimed at understanding the kind of inflation where prices rose at an average rate 2.5 to 3 percent slower than wages. We were also experimenting with a wage-price program with some success. The price explosion that came from food and fuel made all that research and political effort that was aimed at the annoying first few points of inflation look rather trivial and impotent. But when the dust has settled, it is the kind of work we should get back to.

Monetary Growth and the Long-Run Rate of Inflation

By Richard T. Selden*

Arthur Okun invited me to join this panel, one presumes, to present a monetarist view of inflation. This I shall do. In my judgment the current worldwide inflation, like virtually all of its predecessors throughout history, stems from excessive expansion of the money supply. I believe that this could have been prevented, at least in the United States, without either loss of jobs in the long run or danger of financial collapse. The only hope of restoring reasonable price stability is to cut monetary growth to the rate of growth of full employment output and to hold it there indefinitely. Implementing such a policy will not be easy but it is our only hope.

Most recent literature on inflation suffers from two serious faults: it fails to examine the problem in a sufficiently long-term perspective and it takes inadequate account of delays between causal disturbances and inflationary consequences. On the first point, most writers on inflation are preoccupied with ephemeral symptoms or temporary disturbances. Events such as a rise in the minimum wage, a poor corn crop, the formation of an international oil cartel, devaluation of the dollar vis-à-vis other major currencies—the list could on and on—are capable of moving the inflation indicators upward, at least temporarily. However, such alleged causes of inflation operate only sporadically; they cannot explain a sustained rise in prices. One could write a history of inflation that would explain it as a series of unrelated accidents. As social scientists we should eschew such "ad hocery." One way of avoiding this trap is to focus on longer-term price movements rather than on month-to-month or quarter-to-quarter changes. I have concluded that the best time frame for the analysis of inflation is three- or four-year periods. Evidence of this sort is introduced below.

On the second point, the notion that there are long lags in the effects of policy measures is now widely accepted by economists, and even by some policymakers. Nevertheless, I believe that lags in the influence of monetary disturbances on the rate of inflation are much longer than most people believe them to be. For the United States they appear to be three years or more. If I have read the evidence correctly, the major determinant of price trends in 1973–74 is the monetary policies that were followed during the period 1968–71. By the same token, errors in monetary policy during 1971–74 will assure the continuation of rapid inflation into the year 1978.

I.

Let me briefly outline the sort of inflation model that I have in mind. First, I assume that the money supply is subject to close Federal Reserve control over periods of several months. Second, I hold that the monetary growth rate is the major determinant of growth in nominal income. However, this linkage does not work instantaneously. The immediate consequence of an acceleration of monetary growth is a decline in the velocity of circulation. Full recovery of velocity to its normal level requires from four to eight quarters, and accordingly price effects are muted during this adjustment period.

---

* University of Virginia.

Third, like many other economists, I assume that output responds to monetary disturbances faster than prices. The division of changes in nominal income between output and prices in the short run varies directly with the size of the income gap. Finally, I believe that monetary disturbances have no lasting impact on output; the entire effect is on prices.

There is nothing novel about such an inflation model. Broadly similar models have recently been presented by several economists, including Leonall C. Andersen and Denis S. Karnosky, Phillip Cagan, David Laidler, and Bennett T. McCallum. An essential feature is that the influence of money on prices will be spread over long time spans if the adjustments of income to money and of prices to income changes are protracted. Examination of *U.S.* data suggests strongly to me that both conditions hold.

But leaving econometric results aside, is it plausible that prices lag several years behind money? I believe that it is. The case has been stated far better than I can possibly do in these brief remarks in an excellent study by Cagan, *The Hydra-Headed Monster*. For a variety of reasons the impact of aggregate demand on prices works primarily through effects on costs, including notably wage costs. Many of these costs adjust only at infrequent intervals. They are probably influenced in many instances by price anticipations, which appear to be based partly on past price experience. And many prices are controlled or at least influenced by public regulatory bodies.

Given the fact (as I regard it) that the influence of demand on prices is spread over substantial time periods, it is not surprising that in a given quarter nonmonetary factors will dominate the less dramatic but more persistent monetary factor. It is fallacious, however, to infer from this that money does not matter as far as inflation is concerned.

II.

This is not the place to bring forth substantial empirical evidence in favor of the monetarist view of inflation. However, I hope I will be forgiven for introducing one table that summarizes some crude but suggestive evidence based on regressions of quarterly series of three- and four-year percent changes in prices and money. I call such changes the ***long-run*** rates of inflation and monetary growth. The results are displayed in Table 1.

For the United States during the period 1958–I to 1971–II I have found a strong association between money and prices even without lagging money, regardless of the money measure used. However, R^2 was maximized (at .86) when the money variable was lagged fourteen quarters. There was not much difference among the results obtained in the twelve- to fifteen-quarter lag cases. However, R^2 declined steadily as the lag was shortened and when leads were introduced. Addition of the long-run output growth rate—without leads or lags—as a secondary explanatory variable lifted R^2 to .91. The coefficient on output was negative, as expected.

After pondering these *U.S.* results I decided to look at other industrialized economies. Guided mainly by the ease with which reliable data could be put together, I regressed long-run inflation rates on monetary growth in Belgium, Canada, the Netherlands, and Sweden. The results were remarkably consistent with those for the United States. In each instance there was a close association between money and prices, and in each instance the best fit was obtained when the money variable was lagged two years or more. For all of these countries it was

TABLE 1—MONETARY GROWTH AND THE LONG-RUN RATE OF INFLATION, FIVE COUNTRIES

Country	Period	Variables	Lags (qtrs.)	Constant	Ind. Variables	R^2
United States	58-I—71-II	P, M (3-yr. % ch.)	14	0.01	1.04	.86
				(0.02)	(17.56)	
Belgium	58-I—70-IV	P, M (4-yr. % ch.)	8	− 0.02	0.47	.87
				(− 0.13)	(18.21)	
Canada	58-I—71-II	P, M (3-yr. % ch.)	8	0.70	0.23	.83
				(5.53)	(15.66)	
Netherlands	58-I—74-II	P, M (4-yr. % ch.)	8	0.28	0.52	.70
				(0.75)	(12.17)	
Sweden	58-I—71-IV	P, M (4-yr. % ch.)	15	2.53	0.24	.59
				(13.73)	(8.88)	

Notes: Figures in parentheses are t statistics. For the United States, P refers to the *GNP* deflator and M is M-1 (without seasonal adjustment). For other countries, P refers to the consumer price index and M to narrowly defined money.

necessary to work with changes over time spans of three or four years in order to get a proper appreciation of the role being played by money.

How can we explain the differences in R^2 among the countries included in the table? The Netherlands and Sweden, of course, pioneered in the use of incomes policies. Repeatedly they have adopted price freezes and wage restraint as means of coping with inflationary pressures. Even though these episodes of direct controls usually have not lasted very long, they have distorted the normal relationships between money and prices, thereby reducing R^2. In Belgium, Canada, and the United States, on the other hand, the impact of controls has been minimal, up to the early 1970's.

Why lags are longer in the equations for Sweden and the United States is not easy to explain. Both countries tend to follow three-year patterns in collective bargaining, in contrast to Belgium and the Netherlands; but so does Canada. The greater international exposure of Belgium and the Netherlands than of Sweden, and of Canada than the United States, may partly explain these differences.

III.

The behavior of the long-run inflation rate in the United States since mid-1971 is particularly instructive. The long upward march of inflation was halted abruptly after the second quarter of 1971. For the rest of 1971 and throughout 1972 the inflation rate fell quite sharply. On the other hand, the *expected* inflation rate on the basis of prior monetary growth continued to rise. By early 1973, when Phase II gave way to the more flexible Phase III, the discrepancy between actual and expected inflation was huge. The tight controls of Phase II probably would have disintegrated soon if the Nixon Administration had not ended them. 1973 has been called the year of infamy as far as price developments are concerned. Such dramatic language is beside the point. In my view, 1973 was simply a year of "snapback" to inflation levels that were consistent with previous monetary growth. While this adjustment was taking place, the short-run inflation rate soared far above the long-run steady-state rate of about 7 percent per year. Hence, we got double-digit inflation without ever having experienced double-digit monetary growth. But, of course, this hardly refutes the monetarist view.

By the third quarter of 1974, the snapback of prices had been completed. Although there may be some overshooting of actual long-run inflation rates above predicted rates for another quarter or two, I believe that a more likely outcome is a sudden drop of the actual long-run rate to the 6–8 percent per year zone. I expect this kind of a drop in the first half of 1975. However, the excessive monetary growth that has prevailed during much of 1971–74 almost ensures that President Ford's announced goal of whipping inflation by Independence Day, 1976, will not be even closely approximated. The equations in the table suggest that inflation will stay in the 6–8 percent per year range at least until the beginning of 1978. What happens after that depends on Federal Reserve policy in the months ahead. If we are to enjoy price stability in 1978, we must resist the temptation to administer massive monetary stimulation now.

IV.

To conclude these comments, I would like to consider briefly some objections. One frequent objection to monetarism is that it ignores the existence of very substantial reverse causality: monetary growth may be a function of the rate of inflation. My brief reply is that I am not at all persuaded by this criticism, given the long lags previously alluded to. A second objection is that insufficient attention has been paid to the *global* aspect of current inflation. This I regard as largely a consequence of inadequately flexible exchange rates—certainly prior to August 1971 and even since that fateful date. If the United States inflates and other countries tend to support the dollar, then these countries must expect to import inflation from us.

A third objection is more substantial. All of the equations in Table I (aside from the United States equation) underpredict the extent of inflation since 1971. This may reflect failure to measure money correctly through omission of Eurodollar balances. A more likely explanation is that the failure to include output as an explanatory variable is responsible. The sizable decline in those countries' incomes in the wake of the oil crisis may have exerted a significant one-time upward thrust on inflation rates.

Clearly a great deal more work must be done before we can be confident that we fully understand worldwide inflation. In the meantime, policy decisions must be made, for good or for ill. I cast my vote for continued monetary expansion in the 3–4 percent per year range, coupled with a modest tax cut and federal expenditure restraint. Such a policy will not work miracles but it has a good chance of restoring both price stability and full employment by 1978. In the present state of the art we cannot hope to do better.

REFERENCES

L. C. Andersen and D. S. Karnosky, "A Monetary Interpretation of Inflation," paper prepared for Conference on Price Behavior, Nat. Bur. of Econ. Res., Nov. 21, 1974.

P. Cagan, *The Hydra-Headed Monster*, The Amer. Enterprise Instit. for Public Pol. Res., Washington 1974.

D. Laidler, "The 1974 Report of the President's Council of Economic Advisers: The Control of Inflation and the Future of the International Monetary System," *Amer. Econ. Rev.*, Sept. 1974, *64*, 535–44.

B. T. McCallum, "A Small-Scale Neoclassical Macroeconomic Model: Estimates and Tests for Seven Economies," unpublished manuscript.

Interpreting Recent World Inflation

By RONALD L. TEIGEN*

The measured price levels and money wage rates of most industrialized countries began to increase rather rapidly beginning about the mid-1960's, after having risen quite slowly for the previous decade or more. Evidence has been published indicating that the dispersion of individual country inflation rates about the across-country mean declined during this period. Findings along these lines for recent years are presented in the recent Organisation for Economic Co-Operation and Development (*OECD*) special report on the international transmission of inflation, for example. They are generally accepted as fact, probably because theory states that, in a world of trading under rigidly fixed exchange rates, individual country inflation rates must converge to a uniform world rate in the long run. Beginning from the premise that national inflation rates recently have been converging at a noticeable rate, it is perhaps a natural next step to try to identify the dominant causal force which is at work. For instance, the *OECD*, although claiming to hold an eclectic view, takes the position that convergence is due most immediately to the consequences of increasing economic interdependence among trading nations as exemplified, e.g., by the growth in the foreign sector relative to the domestic economy which has been characteristic of many countries. It assigns special importance to the direct effects of import and export prices on domestic costs and prices, suggesting in effect that a type of cost-push process is becoming increasingly important as the world trading system becomes more and more unified.

A popular alternative view, which does not depend on growing interdependence, is that the quickening and converging of national inflation rates are due to a rapid increase in world liquidity over the past few years, with both the accelerating expansion of domestic money supplies in the United States and elsewhere and the massive United States balance of payments deficits at the turn of the decade and into the 1970's being the primary driving forces, but with the introduction of Special Drawing Rights and the development of the Eurocurrency market also playing a significant role. Still others see such factors as demonstration effects, increasing labor militancy, or decisions by stabilization authorities to foster full employment rather than stable prices as the ***primum mobile***.

Whatever the particular viewpoint, the premise that inflation rates among countries are not only moving in the same direction but are noticeably converging toward a uniform rate undoubtedly has encouraged the tendency to identify one set of forces or another as primal. In what follows, I shall present some evidence which casts doubt on the proposition that national inflation rates have been converging in recent years. In doing so, it is not my intention to deny the theory that these rates eventually must converge in a fixed exchange rate world, nor do I reject the proposition that trading countries are becoming more interdependent. What I shall argue is that, since the evidence on convergence is mixed at best, it is preferable to interpret recent experience from an eclectic viewpoint, under which several different causal phenomena probably have been of some importance. As a first step, then, I want to present some evidence on movements in inflation rates and their dispersion.

---

* Professor of economics, University of Michigan.

The *OECD* evidence on convergence of national inflation rates in the special report mentioned above is summarized in a figure (*OECD*, Chart R, p. 82) in which the coefficients of variation of the *GNP* deflators for seventeen *OECD* countries, as well as for thirteen Northern European countries, are plotted annually over the period 1956–73. A trend line with a pronounced negative slope is fitted through these observations, and it indicates that, if past and current tendencies were to continue, the coefficient of variation for the *GNP* deflators of these countries would be zero by about 1982. This downward-sloping trend line is the basis for the *OECD*'s view that world inflation rates are converging.

In considering this evidence, it should be recalled that the coefficient of variation, being the ratio of the standard deviation to the mean, will decline either because of decreasing absolute dispersion or a rising mean. If a reduction over time in the coefficient of variation caused by a rising mean inflation rate and a constant dispersion of individual country rates is interpreted as signifying a convergence of these rates, it is implied that this dispersion, as measured by the standard deviation, should normally change proportionately to the mean. It is not clear why this would be expected in the case of national inflation rates, especially in a world in which exchange rate changes, exogenous capital movements, and other factors interfere with the adjustments which would occur under a hypothetical fixed exchange rate system. Therefore an absolute measure of dispersion such as the standard deviation may be more appropriate in this context than the coefficient of variation. In any case, there is some evidence that the dispersion of inflation rates may have begun to increase in recent years even when measured by the coefficient of variation. It is to this evidence that I now turn.

I have not been able to review the year-by-year experiences of all of the *OECD* countries. However, the data on compound growth rates of the consumer price index (*CPI*) and the *GNP* deflator which are calculated and published by the Federal Reserve Bank of St. Louis in its report entitled *Rates of Change in Economic Data for Ten Industrial Countries* are very convenient for my purpose. The countries included are Belgium, Canada, France, Germany, Italy, Japan, the Netherlands, Switzerland, the United Kingdom, and the United States; together they account for 84 percent of the imports and 85 percent of the exports of all *OECD* countries (as of 1971). I have calculated the means and standard deviations for the rates of change of the *CPI* and the *GNP* deflator over the three subperiods 1954–65, 1965–71, and 1971–73. The points of demarcation are somewhat arbitrary, but are meant to correspond roughly to the beginning of the acceleration of the Vietnam conflict (1965) and the breakdown of the fixed exchange rate system (1971). The results are reported in Table 1.

Both of these indexes record an acceleration in the rate of price increase after 1965 and a considerable further speedup following the breakdown of the Bretton Woods system in 1971. The numbers are slightly different, with the *GNP* deflator indicating higher growth rates than the *CPI*, but the patterns are the same. However, the patterns of dispersion for the two indexes are somewhat different, whether dispersion is measured by the standard deviation alone or by the coefficient of variation. Using the *CPI*, the standard deviation is constant for the first two subperiods and then increases during 1971–73. The coefficient of varia-

TABLE 1—MEAN ANNUALLY COMPOUNDED GROWTH RATES OF THE CONSUMER PRICE INDEX AND IMPLICIT GNP DEFLATOR, AND THEIR DISPERSIONS, FOR TEN INDUSTRIAL COUNTRIES, 1954–73[a]

	Consumer Price Index	Coefficient of Variation	Implicit GNP Deflator	Coefficient of Variation
1945–65	(In Percent)		(In Percent)	
Mean	2.7	0.33	3.3	0.27
Standard deviation	0.9		0.9	
1965–71				
Mean	4.3	0.21	4.7	0.13
Standard deviation	0.9		0.6	
1971–73				
Mean	7.0	0.16	7.1	0.18
Standard deviation	1.1		1.3	

[a] The data on the implicit *GNP* deflator for Japan begin in 1955.

Source: Federal Reserve Bank of St. Louis, *Rates of Change in Economic Data for Ten Industrial Countries*, Oct. 1974.

tion falls in the latter two subperiods compared to the first, due to the sharply increasing mean inflation rate. Using the implicit *GNP* deflator—the same index employed by the *OECD* in the aforementioned study—the standard deviation falls by one-third in 1965–71 while the mean rises by more than one-third; this results in a halving of the coefficient of variation. However, the standard deviation doubles in 1971–73 as compared with 1965–71. Even though the mean inflation rate also rises sharply in this subperiod, the coefficient of variation rises considerably as compared to 1965–71.

At least for these ten important industrial countries, then, the evidence suggests that national inflation rates are not clearly converging, but rather appear to have become more diffuse in recent years. While some divergence might be expected for 1971–73 due to structural changes in the world monetary system, it is not clear that national rates were converging even in earlier years. As noted above, this finding tends to weaken single-cause views of the world inflationary process. Other evidence, such as there is, also is unfavorable to this approach. For example, studying the experience of seven countries (six of which are included here) over the period 1955–71, W. D. Nordhaus found that neither simple monetarist nor simple Phillips curve explanations of wage and price changes were adequate.

If single causes cannot satisfactorily explain recent experience, we shall have to rely on more *ad hoc* combinations of forces, events, and institutions. In this spirit, let us first consider the period 1971–73. I believe that much of the observed acceleration of inflation rates in these years was due to two causes. One, which is difficult to quantify, is the increased freedom to pursue domestic goals made possible for trading countries by the collapse of the Bretton Woods system. The other is the food shortage and consequent sharp rise in food prices beginning in 1972–73 (the oil price increases of late 1973 probably are not reflected in these data). Among the ten countries studied here, the rate of *CPI* increase rose in 1972–73 compared to 1971–72 by 20 percent or more in nine countries and by 30 percent or more in five. In a world characterized by downward price inflexibility, and in which high employment is an important goal, it is likely that the relative prices of other goods were lowered in response to the food shortage, and employment supported by accommodative policy and the acceptance of higher inflation rates.

Similar considerations probably apply to many countries in the period prior to 1971. Stabilization authorities have become more knowledgeable in managing demand, and at the same time, maintaining a high level of activity has become relatively more important as a policy goal. As a consequence, industrialized countries are operating closer to full employment more of the time than previously. This suggests both the acceptance of a higher average inflation rate and less possibility for error on the expansionary side without inflationary repercussions.

The other set of factors whose importance I want to emphasize is monetary in nature. Although I do not agree with the monetarist view that increases in world liquidity are both necessary and sufficient to generate a rising rate of world inflation, I certainly believe that such increases have in fact played an important role in recent years. Among the ten countries examined, the unweighted annual average compound growth rate of their money supplies was 8.0 percent in 1954–65, 9.6 percent in 1965–71, and 13.7 percent in 1971–73. Certainly this must have been expansionary; the problem lies in determining what part of this growth represented policy initiatives and what part was accommodative in the sense of a policy response to external conditions, designed to maintain employment or defend some other goal.

Whatever their origin, monetary impulses propagated through the trading world by balance of payment changes can create special problems for countries with underdeveloped financial markets and institutions in which it is difficult for the authorities to neutralize the effects of such changes on the domestic monetary base. Open market operations typically cannot be used in such economies, and other devices such as reserve requirement changes have their own substantial limitations. Consequently, capital flows tend to have immediate and substantial repercussions on domestic liquidity. Those originating in the United States balance of payment deficits in the pre–1971 period must have resulted in substantial increases in liquidity, and consequently in expansion and in upward pressure on prices, in many of these countries.

REFERENCES

W. D. Nordhaus, "The Worldwide Wage Explosion," *Brookings Paper Econ. Act.*, No. 2, 1972, 431–64.

Federal Reserve Bank of St. Louis, *Rates of Change in Economic Data for Ten Industrial Countries*, St. Louis, Oct. 1974.

Organisation for Economic Co-Operation and Development, "The International Transmission of Inflation," *Econ. Outlook*, July 1973, 81–96.

DOMESTIC ADJUSTMENT TO BALANCE OF PAYMENTS DISEQUILIBRIUM

The Payments Adjustment Process and the Exchange Rate Regime: What Have We Learned?

By MARINA v. N. WHITMAN*

In taking stock of what we have learned about domestic adjustment to exchange rate change since the end of the Bretton Woods system, the first step is to try and define the nature of the exchange rate regime we are talking about. Certainly, the exchange rate situation of the past few years is a far cry from the system of freely flexible rates so beloved of textbook writers. Rather, it is a state of transition, without clearly defined rules of behavior, characterized by "probably the widest combination of exchange systems ever with the exception perhaps of the mid-1930's." As of mid–1973, an official of the International Monetary Fund (*IMF*) noted:[1]

> that at present we have 10 countries floating with different degrees of intervention; we have 8 countries floating as a bloc, maintaining a very close relationship among themselves but with no obligation whatsoever to defend a rate with respect to the dollar; and we have 24 countries that have pegged their currencies with respect to the dollar; 14 with respect to the French franc; and 11 with respect to the pound; and we have 53 other countries that have either declared central rates or par values which they defend by intervening in any of the three above-mentioned currencies. [p. 107]

I. The Nature of the Present System

As this brief summary suggests, the present situation is one characterized by active and to a considerable extent uncoordinated behavior on the part of national central banks. Many countries behave as though they have explicit current account targets, and therefore implicit exchange rate targets, rather than being willing to allow exchange rates to move freely to clear the foreign exchange markets. The present lack of coordination is evidenced not by the direct intervention of countries in the foreign exchange markets, regarding which there has apparently been considerable consultation, but rather by what might be termed their indirect intervention: central bank or government activities which not only have a significant impact on the foreign exchange markets but which are undertaken for the primary purpose of modifying or restricting the behavior of those markets.

One form of indirect intervention by which countries may seek to avoid or reduce reserve losses or, alternatively, depreciation of their currencies is by foreign (as opposed to domestic) borrowing by governmental or government-related authorities. France, Italy, the United Kingdom, and Japan were among a substantial number of countries which utilized large-scale foreign borrowing by public-sector institutions, particularly in the Eurodollar market, to strengthen their payments positions late in 1973 and early in 1974.

---

* Distinguished Public Service Professor of Economics, University of Pittsburgh.

[1] Statement of Ricardo H. Arriazu, Alternative Executive Director, *IMF*, in U.S. Congress (93–1) p. 107. See also, Council of Economic Advisers, Table 53, p. 197.

The most common and time-honored method of modifying exchange rate or reserve positions is, of course, some form of direct control over international transactions. The period since August 15, 1971 has witnessed a widespread escalation of controls for balance of payments purposes, imposed primarily by surplus countries to limit capital inflows and thus prevent or, more often, hold down the appreciation of their currencies. The fact that such controls were relatively ineffective in stemming the tide of mobile funds did not prevent the surplus countries from applying them with progressive tightness and in increasingly diverse forms.

Even after the move to a system of what was termed "generalized floating" early in 1973, the use of direct restrictions to modify external positions continued.[2] The widespread relaxation of import restrictions and the imposition of taxes or limitations on exports during 1973 were, with a few exceptions, directed at the alleviation of domestic scarcities and inflationary pressures rather than at external targets. But toward the end of that year, following the sudden developments affecting the price and availability of petroleum imports, this trend among the developed countries toward liberalization of import restrictions tapered off. In a few countries with severe payments problems the liberalizing moves were in fact reversed; the most dramatic example of this was the imposition by Italy in May 1974 of a 50 percent deposit requirement on imports, including imports from its partner countries in the Common Market.

Despite a few such widely noted exceptions, direct controls for balance of payments or exchange rate purposes have continued to be focused primarily on the capital rather than the current account. Efforts to curb capital inflows were stepped up in many industrialized countries early in 1973, often accompanied by measures to liberalize capital outflows. But by the latter part of 1973, after the advent of the petroleum problems, many developed countries moved to liberalize capital inflows and, in a few cases, even reimposed curbs on capital outflows. There was one major exception: the United States programs to restrict capital exports had been liberalized in 1973 and, in January 1974, encouraged by the substantial strengthening of the dollar over the previous six months and by the safety valve against speculative disasters provided by floating rates, the *U.S.* government abolished all three of these programs. Nonetheless, despite some major exceptions, the early experience with greater flexibility of exchange rates suggests that, contrary to a priori expectations, it has been accompanied by an increased reliance on direct controls aimed at the modification of countries' external positions.

The "floating rate" system of the recent past has clearly been heavily managed, by both direct and indirect intervention; it has also been characterized by substantial fluctuations in exchange rates. In the eighteen months following the Smithsonian Agreement, the trade-weighted effective exchange rate of the dollar against thirteen major currencies fell by some 12 percent; in the next six months it rose again by an almost equal percentage and then fell again by some eight percentage points between January and May of 1974, to a level nearly 20 percent below pre-June 1970 parities, before rising again some five percentage points or so (Morgan Guaranty Trust Co.). Several other major countries experienced even larger changes in their effective rates of exchange (in May 1974 Germany's effective rate was 22 percent above and Italy's 25 percent below their respective pre-June 1970 parities) and fluctuations in certain important bilateral rates (such as the dollar-mark and dollar-yen rates) were larger still. Movements of several percentage points in the dollar-mark rate took place within a single week—occasionally, even within a single day.

---

[2] For surveys of payments restrictions in 1973, see Council of Economic Advisers, Table 54, pp. 198–99; International Monetary Fund *Annual Report*, pp. 44–46; and International Monetary Fund *Survey*, Sept. 2, 1974, pp. 274–76.

In a period characterized by sudden, dramatic changes in economic conditions affecting countries' external positions and by high and widely divergent rates of inflation, it is perhaps not surprising that movements in exchange rates were large, sharp and, on occasion, dramatically reversed. Nonetheless, these fluctuations do appear to have been somewhat larger than required for the elimination of latent flow disequilibria—balance of payments disequilibria in the usual sense. Some scholars have taken this as evidence that the exchange markets are inherently unstable. (See, for example, C. P. Kindleberger.) Others see it, however, as a transitional phenomenon, a temporary divergence between stock equilibrium and flow equilibrium resulting from a prolonged period of flow disequilibrium (as, for example, W. S. Salant). In this view, once the inherited dollar overhang is liquidated and the desired shift in the composition of reserve assets, both official and private, is completed, a major cause of the volatility in the valuation of the dollar vis-à-vis other currencies should disappear.[3]

II. Exchange Rate Change and the Stability of Foreign Exchange Markets

The fundamental question to which we seek answer from recent experience is, of course, whether greater flexibility of exchange rates is indeed an effective means of reducing or eliminating payments disequilibrium, thus alleviating the burden of adjustment which must otherwise be borne either by internal measures or by direct restrictions on international transactions. Because of limitations of time and space, I shall focus here on the experience of the United States, asking whether the effective devaluation of the *U.S.* dollar since mid-August 1971 has or has not served to reduce the deficit in the *U.S.* payments balance.

One difficulty with discussing adjustment to balance of payments disequilibria in the framework of a hybrid system of "managed floating"—in which disequilibrium is dealt with partly by exchange rate changes and partly by reserve changes—is that the system no longer always gives unambiguous signals indicating which side of equilibrium we are on. In such a situation, when at any given moment some countries are likely to be intervening in the exchange markets and others are not, it is possible for the dollar to be depreciating at the same time that the United States has an official settlements surplus, or vice versa. Although it is still generally possible to garner consensus views regarding a country's underlying external position, data on payments positions and changes in reserve balances no longer serve as unambiguous indicators in this regard.

Whereas during 1972 and the early part of 1973 the prevailing mood with respect to the *U.S.* balance of payments was one of elasticity pessimism (as evidenced by the further depreciation of the dollar in exchange markets following the official devaluation of February 1973), during the course of 1973 the tide began to turn, and the conviction grew that the Smithsonian devaluation was indeed having a favorable impact on the external position of the United States. Such systematic econometric evidence as is beginning to be available tends to confirm the conclusions derived from more casual empiricism, that the effective depreciation of the dollar has indeed been "successful."[4]

---

[3] Indeed, the demand for dollars to make oil payments appears already to have reduced substantially the magnitude of the "dollar overhang" problem in the eyes of other industrialized countries.

The econometric evidence available so far is scanty, piecemeal, and far from definitive. In particular, the studies which have been completed so far suffer from three major weaknesses as predictors of the impact of recent changes in the effective exchange rate of the dollar. First, they are carried out within the partial equilibrium framework of the elasticities approach, explicitly ignoring the induced effects of devaluation on domestic and foreign incomes and price levels. Such *ceteris paribus* estimates must be regarded as upper bounds for the *mutatis mutandis* effects of depreciation. Second, these studies encompass the effects on the merchandise trade balance only; virtually no econometric analysis of the impact on other elements of the current account or the long-term capital account is yet available. (An exception is S. Y. Kwack, 1974b.) Finally, these studies are also partial in the sense that the values of the underlying price elasticities for various components of exports and imports are assumed rather than estimated directly. That is, they utilize some combination of a priori reasoning and estimates derived from preexistent econometric studies to generate either point estimates or range estimates of price elasticities on which to base the analysis of the impact of depreciation on the *U.S.* trade balance. To my knowledge, no analysis of the 1971–73 period has generated its component estimates of price elasticities from scratch.

Given these inherent limitations, it is nonetheless encouraging that the various studies (W. Branson, P. B. Clark, and Kwack 1974b) are in broad agreement in concluding that a substantial improvement in the *U.S.* trade balance had already taken place by the end of 1973 as a result of the dollar depreciation, and that an even larger total improvement could be expected ultimately to result.[5] The various studies vary substantially in their estimates of the magnitude of the depreciation effect and they are noncomparable in many ways, including the models employed, the time periods of both observation and prediction, and whether or not the analysis related to the Smithsonian devaluation alone or to the total change in the effective exchange rate of the dollar over the 1971–73 period. Nonetheless, all generate estimates of a total favorable impact on the *U.S.* trade balance of several billions of dollars, although no estimate is as large as Secretary of the Treasury John Connally's original target of a $13 billion improvement in the merchandise trade balance.

The various studies of the impact of dollar depreciation on the *U.S.* trade balance differ in their estimates not only of the magnitude of the effect, but also of its timing. Estimates of the lag in the response vary from relatively short—implying that more than 50 percent of the total effect on export and import quantities takes place within one year (Kwack 1974b)—to rather long—suggesting that more than three years are required for half of the ultimate impact to have occurred (H. B. Junz and R. R. Rhomberg). But, despite these differences, the studies tend in general to confirm empirically the existence of a phenomenon which had been increasingly anticipated on a priori grounds,[6] the so-called J-curve effect.[7] The term J-curve refers to the existence of a nontrivial period of time during which the negative price effects of a depreciation dominate the positive quantity effects, implying that the initial effects of an exchange rate change on the balance of trade are perverse even when the elasticity conditions for long-run stability are satisfied.

---

[4] R. I. McKinnon (pp. 24–26) cites an econometric analysis of the impact of exchange rate change in six leading industrialized countries over the period 1968–74 to support his view that, from this broader perspective, the results are mixed. The study referred to is Wolfgang Kasper's.

[5] These are obviously *ceteris paribus* predictions, which leave out of account the substantial deterioration in the 1974 trade balance stemming from the fourfold increase in petroleum prices.

As regards the other components of the *U.S.* basic balance—the current account items other than merchandise trade and long-term capital flows—very little empirical analysis is yet available for the 1971–73 period. In terms of raw annual data, a steady improvement in the balance on long-term private capital flows from 1971 through 1973 is evident, without the dip in 1972 which characterizes the merchandise trade and current account positions. There are two reasons why one might expect a priori that a change in the exchange rate (as distinguished from the expectation of a future change in the exchange rate) will affect the net flow of long-term capital relatively quickly. One is the termination or reversal after a devaluation of capital outflows which had taken place previously in anticipation of that devaluation, a phenomenon which one would expect to be more significant under a system of adjustable pegs than in situations characterized by widespread floating. The second is the so-called portfolio-rebalancing effect, which arises from the assumption of portfolio theory that investors want to hold a particular optimum mix of assets in their portfolios and will attempt to maintain these proportions, denominated in their own domestic currencies, in the face of valuation changes arising from exchange rate changes. (See D. E. Logue and T. D. Willett.) Over the longer run, the improved competitive position of domestic export- and import-competing industries resulting from depreciation can be expected to attract increased investment from abroad.

Although empirical evidence regarding the role of these various effects in the recent period is as yet lacking, it seems on a priori grounds that, unless the portfolio-rebalancing effect is significant, changes in long-term capital flows, which are largely nonspeculative in origin, are likely to take place only gradually and with a significant lag following an exchange rate change. If this is so, it raises a question about the dynamic stability of foreign exchange markets in a world where markets clear rapidly and in which the initial response of the trade balance to exchange rate changes may be perverse (A. J. C. Britton). Actually, the question is a double one: Are exchange markets stable in the absence of government intervention and, if so, what makes them so?

Despite substantial fluctuations, the exchange markets for the dollar and other major currencies do appear to have been dynamically, though cyclically, stable in the period since mid-1971. That is, although fluctuations in rates may well have been larger than would have been required to eliminate underlying disequilibria, the major rate movements appear to have been in the "right" direction, that is, appropriate to actual or anticipated changes in basic economic conditions. Certainly rates did not appear to be in danger of shooting off toward zero or infinity. Nor could government intervention account fully for the existence of dynamic stability; such intervention, although substantial, has not been continuous or systematic enough to explain market behavior.[8] Rather, it would appear that private speculation, although inadequate in magnitude and speed of response, has indeed performed its traditional function of being stabilizing in the broad, despite widespread concern about "bandwagon" behavior unrelated to fundamental economic conditions.

---

[6] See S. P. Magee for a systematic discussion of the conditions required for a J-curve to exist.

[7] Again, Kwack (1974b) is an exception.

A pattern of speculative behavior which magnifies the amplitude of currency swings, requiring price changes which overshoot equilibrium in order to call forth equilibrating movements of funds, appears to be due to a paucity of continuous stabilizing speculation, resulting in the market thinness just described.[9] Although, again, the evidence at this stage is primarily anecdotal, there seems to be a widespread feeling that, despite a few widely publicized examples to the contrary, large *U.S.* banks have become increasingly cautious about assuming net positions in foreign currencies, either spot or forward. Even more marked is an apparent tendency for multinational corporations to overhedge against foreign exchange risk, owing in part to the current uncertainty among accountants as to the correct method for analyzing "translation costs," the changes in the domestic-currency valuation of foreign assets and liabilities resulting from a change in the exchange rate between the currency of the parent country and that of the country in which a subsidiary is domiciled.

Behavior of the sort just described is what one might expect during a learning period, when speculators' views regarding long-run equilibrium values are weakly held and substantial stimulus is therefore required to make them act on them. Certainly such learning behavior is consistent with the cyclical stability which appears to have characterized foreign exchange markets since the demise of Bretton Woods. However, much more systematic analysis of the behavior of exchange markets during this period, and of the roles played respectively by private speculation and by central bank intervention, is urgently required. Furthermore, sufficient time must elapse and the exchange rate system must settle down sufficiently to permit a distinction between temporary behavior associated with rapid transition and the more permanent behavior patterns associated with the new exchange rate regime before views regarding market stability can be held with reasonable certainty.

III. Feedbacks to the Domestic Price Level

The empirical analysis cited so far has focused primarily on changes in relative prices, abstracting from induced changes in domestic and foreign incomes and price levels. But there is a growing school of thought which holds that such analysis is fundamentally irrelevant; that the primary impact of exchange rate change on the balance of payments comes not through changes in relative prices of commodities, but rather through the direct effects on expenditure of the resulting change in the real value of money balances. An improvement in the balance of payments of a depreciating country, according to this "global monetarist" view, is a transitory phenomenon, stemming from the reduction in the real value of domestic cash balances as a result of the price increases associated with depreciation.[10] This fall in real money balances, if it is not offset by the monetary authorities, produces an excess demand for money and thus a temporary drop in expenditure relative to income. But, once domestic residents have built their real cash balances back up to the desired level through an increased payments surplus (or reduced deficit), the balance of payments will revert to its previous level. A further implication of this view is that any initial change in relative prices stemming from an exchange rate change will ultimately be offset by a combined increase in the price level of the depreciating country and decrease in the price level of the appreciating country.[11]

---

[8] R. A. Debs (p. 87) notes that official intervention in exchange markets between March 1973 and March 1974 exceeded $35 billion; my own rough calculations based on monthly changes in reserves suggest that intervention by G-10 countries totalled more than $35 billion during the calendar year 1973 and about $60 billion between the end of 1972 and October 1974. J. Williamson concludes from his statistical analysis of the experience of twenty countries that reserve use may actually have increased, and certainly has not decreased, under flexible as compared with fixed rates.

[9] McKinnon (pp. 17–21) also concludes that stabilizing speculation was lacking in 1973–74, and suggests that "the decline in the United States dollar as stable international money" may be a major reason for its paucity.

The controversy generated by this approach is less over whether the balance of payments effects of exchange rate change are transitory than over how long the transitory period is. The fact that, over long periods of time (e.g., decades), changes in exchange rates tend to offset—or be offset by—changes in relative price levels has long been recognized in the form of the Purchasing Power Parity concept.[12] The question at issue is whether this process of adjustment actually takes place relatively quickly. The global monetarist view is that high elasticities of substitution prevail among countries for most tradeable goods and that, because world markets are today highly integrated, a single price must prevail in all markets for goods which are close substitutes for one another.[13] The implication of this view is that competitive forces will quickly eliminate changes in relative prices stemming from exchange rate changes by offsetting changes in domestic prices.

No full-fledged econometric analysis of the balance of payments impact of recent exchange rate changes which utilizes the macroeconomic approach of the global monetarists has yet come to my attention. However, some empirical work has been focused on one hypothesis closely associated with this approach: the close-substitutes hypothesis.

While some of the empirical investigations of the effect of exchange rate changes on trade balances, discussed earlier, simply assume that certain categories of goods moving in world trade (chiefly primary commodities) conform to the close-substitutes model while other categories (mainly manufactured goods) do not, other studies attempt to evaluate the close-substitutes hypothesis directly. In general, the conclusion is that, at least for manufactured goods, the close-substitutes hypothesis fails to hold for the periods examined and at the levels of aggregation employed; that, in other words, "the relative dollar prices of industrial outputs do seem to change substantially and permanently (over the six-year period

---

[10] Examples of this approach are R. Dornbusch, H. G. Johnson (1969, 1972a, and 1972b) and A. B. Laffer. Whereas the elasticities approach deals with the impact of exchange rate change on the balance of *trade*, the global monetarist approach focuses on the impact on the balance of *payments*.

[11] Kasper concludes that "the evidence suggests that exchange rate changes had a more immediate and more substantial impact on export and import prices than on export and import volumes . . . the dominant impact of exchange rate changes on the prices of foreign trade would also suggest that the exchange rate instrument be assigned to the (domestic) price-level objective, especially in countries with open economies."

[12] See H. J. Gailliot, Laffer, and the references there cited. This may not be true in the short run, however. McKinnon (p. 29) notes that the sharp appreciations of European currencies against the dollar in 1971 and 1973 "seem completely inexplicable to a theorist relying on some notion of purchasing power parity. . . . Changes in exchange rates in 1973 and 1974 cannot be interpreted as better aligning national price levels."

[13] Johnson (1972b, pp. 1560–1561) notes that this close-substitutes model is a common but not an essential assumption of the global monetarist approach.

1968–73) when exchange rates change" (P. Isard, p. 17). Kwack (1974a) estimates that, on the average, a one percentage point devaluation would improve the relative price advantage of the United States by 0.7 percent (with *U.S.* prices rising in the long run by about 0.2 percent and foreign prices falling by about 0.1 percent.) J. R. Artus, looking only at the prices of manufactured exports, estimates that in the United States domestic price changes will ultimately (within two years) offset about 10 percent of an exchange rate change, while for four other major industrialized countries the proportion is likely to be between 30 and 40 percent. Combining these estimates implies, again, that the United States can expect to retain for some time more than half of a change in the effective exchange rate in the form of a relative price advantage for manufactured exports.

The close-substitutes model differs from the conventional view also in its implications regarding the effect of an exchange rate change on the domestic price level or rate of inflation. Under the assumptions of the conventional Keynesian model, that wages and prices are stable in terms of the home currency up to the point of full employment, depreciation of the exchange rate in the presence of unemployment would affect the domestic price level only via the increased prices of imports. Under these assumptions, the ratio of imports to *GNP* would set an upper limit to the impact of an exchange depreciation on the domestic price level. For the United States, whose imports constitute some 7.5 percent of *GNP*, an effective dollar depreciation of 10 percent could not increase the price level, as measured by the *GNP* deflator, by more than three-fourths of 1 percent (this represents an upper bound; because of the relative importance of the United States in world markets, it is most unlikely that the full amount of the depreciation would be passed through in the form of increased dollar prices of imports.) (See Isard, p. 3.) Under the assumptions of the perfectly competitive model, on the other hand, with elasticities of substitution between foreign and domestic goods approaching infinity, the upper bound on the increase in the domestic price level would be the amount of the effective depreciation itself. Actually, because the United States is far too large to be a price-taker in world markets, the dollar depreciation would be divided in some proportion between a rise in price level in the United States and a fall in the price levels of those countries whose currencies appreciated vis-à-vis the dollar.[14]

Although systematic evidence concerning the impact of the dollar depreciation on the acceleration of inflation in the United States is still scarce, two early studies which incorporate at least some of the indirect effects of depreciation on the domestic price level implied by the close-substitutes model have yielded reasonably consistent results. Utilizing regression results based on Phillips curve equations which incorporate an import-price term for the United States and "the rest of the world" over the period 1959–71, Kwack (1974a) estimates that, on the average, the *U.S.* price level as measured by the consumer price index would rise by about 20 percent of the effective depreciation of the dollar, or by two percentage points in the case of an effective devaluation of 10 percent. W. Nordhaus and J. Shoven, who use an input-output approach to estimate the transmission of the price effects of depreciation from the "exogenous" sectors corresponding to the close-substitutes model to the "endogenous" or sheltered sectors of the economy, conclude that the 10 percent effective devaluation of the dollar between November 1972 and August 1973 accounted for between one-fifth and one-sixth of the *U.S.* inflation, as measured by their adjusted wholesale price index, over the same period. Since inflation over that period, in terms of their index, amounted to 11.3 percent, and since their index appears to be closer in concept to the consumer price index than to the wholesale price index as actually measured, their attribution of between 1.9 and 2.3 percentage points of inflation to the effective depreciation of the dollar neatly brackets Kwack's partially a priori estimate of 2 percentage points.

---

[14] J. Salop notes that "the monetary approach most commonly appears as a small country model," in which the domestic price level rises by the full amount of an effective depreciation.

Although the studies just described yield far larger estimates of the impact of depreciation in the domestic price level than did the very crude *ex ante* calculations cited earlier, there is reason to believe that they may still underestimate this impact, by failing to take account of significant indirect effects on prices in industries producing exportable and import-competing goods and, through effects on rates of wage increase, on prices in nontradeable sectors as well. In particular, the improvement in our net balance on goods and services, in real terms, from a deficit of $0.8 billion in the fourth quarter of 1972 to a surplus of $11.6 billion in the fourth quarter of 1973 (in 1958 dollars) represented 38 percent of the total increase in real *U.S. GNP* over the period. At a time when many parts of the economy were clearly straining against capacity limitations (a pressure apparently not fully reflected in the rate of unemployment), the diversion of nearly 40 percent of the increase in real output from domestic to foreign absorption cannot help but have had effects on the domestic price level far more complex and pervasive than those captured in either of the models just described. In reality, as Johnson (1969) has suggested, neither the conventional closed economy model nor the fully integrated model of the monetarists is a realistic one; what is needed is a model which specifies differing degrees of imperfect substitution between domestic and foreign output in different sectors of the economy.[15] Although the studies just described make some significant steps in that direction, the full specification of a disaggregated model of imperfect substitution appears to be still some way off.

IV. The Exchange Rate Regime and World Inflation

The existence of a positive relationship between the depreciation of the dollar and the acceleration of domestic inflation in the United States is beyond dispute; the only questions relate to its magnitude. The relationship between changes in the exchange rate regime since mid-1971 and the dramatic increase in the *world* rate of inflation over the same period is, however, much less clear. Obviously, changes in exchange rates can be expected to push inflation around geographically, accelerating it in depreciating countries and retarding it in appreciating countries. But in formal terms these effects should be symmetrical, canceling each other out and leaving the world rate of inflation unchanged. Any net effect of exchange rate changes on the rate of inflation in the world as a whole requires asymmetries in either market or policy responses.

That such asymmetries should exist is not surprising; clearly there are differences among countries in their elasticities of supply and demand, in other structural characteristics of their economies, and in the responses of their policymakers. The problem is simply that it is by no means clear in which direction these asymmetries operate and, therefore, what the relationship is on balance between different exchange rate regimes and the world rate of inflation. Proponents of fixed rates, on the one hand, argue that a move to flexibility destroys balance of payments discipline as an external constraint on governments' inflationary policies, that the increased uncertainty associated with flexible rates creates an inflationary "flight from money," and that, if wages and prices are inflexible downward and factor mobility is imperfect, the excessive fluctuations or "overshooting" frequently associated with freely flexible rates will have an inflationary ratchet effect on the domestic price level, as well as moving the country's Phillips curve outward in an unfavorable direction.[16] Supporters of flexible rates, on the other hand, note: that in a world of downward inflexibility, the failure of deficit countries to deflate forces surplus countries to inflate against their wills; that the international liquidity creation associated with prolonged payments disequilibrium is bound to be inflationary; and that the large speculative capital flows generated by disequilibrium rates under constant pressure are on balance inflationary, particularly when they flow out of the reserve-center country as ordinary money and into other countries as high-powered money.

---

[15] R. J. Gordon notes that the close-substitutes monetary model is appropriate to the long run, while the closed-economy Phillips-curve approach is applicable to short-run policy analysis.

Despite considerable uncertainty regarding the nature of the asymmetries which govern the relationship between the exchange rate regime and the magnitude of world inflation, there is a persuasive case to be made that the Bretton Woods system, particularly in its final years, represented a particularly inflationary organization of the international monetary system. The four major elements of a one-sided adjustment process, a parity system based on a structurally weakening dollar, destabilizing capital flows, and uncontrolled expansion of international liquidity combined to produce a highly inflationary mixture in which neither the United States nor the other industrialized countries, taken together, felt external constraints on domestic inflationary pressures. (See Johnson 1969 and Emminger.)

At present, the international transmission of inflation is a painful phenomenon in search of a theory; "we have no general coherent theory of world inflation fully applicable to the present world monetary environment characterized by extensive floating" (G. Haberler, p. 4). Nonetheless, recent experience has made at least a few points clear. First, because of systematic differences between the United States and other countries in the relationship between the behavior of tradeable goods prices and that of more generally utilized indices like the consumer price index or the *GNP* deflator, it was possible for the United States to export inflation even when its own rate of inflation was very moderate and below that of other industrialized countries. Second, domestic credit creation may precede or preclude an inflow of reserves in a potential surplus country; such a reserve inflow is a sufficient but not a necessary condition of imported inflation (Johnson 1973 and Emminger).

The twilight years of the Bretton Woods system created widespread awareness that, without exchange rate flexibility, countries other than the key-currency country can exercise little control over their domestic money supplies and price levels. The promise of greater autonomy for domestic monetary policy has been one of the major arguments in favor of greater flexibility of exchange rates. However, the post-Bretton

---

[16] Kindleberger argues that floating rates add to structural inflation in some periods (as at present) and to structural deflation in others (as in 1929–32).

Woods experience has made it clear that the monetary autonomy made possible by flexibility of exchange rates has definite limits. First of all, such monetary autonomy requires not simply *flexibility* of exchange rates but *uncertainty* regarding the future course of exchange rates—uncertainty which at the same time threatens the continued growth of international trade and investment and, many fear, will eventually lead to the fragmentation of world markets. A moving exchange rate, such as a crawling peg, whose future values are known with certainty, allows no more room for monetary independence than does a fixed rate, although it alters the particular configuration of monetary policies that is consistent with the exchange rate regime. Second, while rate flexibility endows a country with greater control over the nominal money supply, to the extent that there are feedbacks from changes in the exchange rate to the domestic price level, the real money supply remains an endogenous variable rather than a policy variable subject to the control of the domestic monetary authorities.

The move to widespread rate flexibility has not provided complete autonomy for national policies in pursuit of domestic economic goals; neither has it abolished problems associated with the supply of international liquidity, either on the world or the individual country level. Because exchange rates are heavily managed, with considerable intervention in foreign exchange markets, one would hardly expect the demand for international reserves to drop to zero. At the same time, it is surprising that, in the words of the 1974 *Annual Report of the IMF* (p. 39), "there is not as yet any statistical evidence of a reduction in the utilization of reserves by countries that have allowed their currencies to float." The shift to widespread floating appears to have halted the explosive growth of world reserves which occurred in 1970–72 without clearly increasing, as one might have expected, the redundancy of the existing reserve stock by reducing the utilization of reserves.

Finally, while the change in the exchange rate regime failed to provide complete autonomy for national economic policies, neither has it spawned a rash of beggar-my-neighbor exchange rate policies. There has been no wave of competitive depreciations, perhaps because countries are alert to the dangerous and ultimately self-defeating characteristics of such behavior, perhaps also because the twin concerns of stagflation have brought about a certain symmetry in the pressures for depreciation to alleviate unemployment and for appreciation to alleviate domestic inflation.

Despite substantial fluctuations in exchange rates, the widely feared decline in international trade and investment has failed to materialize. Although the volume of world trade grew by only about 6 percent in 1971, well below the 1960–1970 average of 9 percent, this slowdown appears to have been related to the worldwide cyclical decline in output growth during 1970–71, rather than to the changes in the exchange rate regime which occurred in the second half of 1971. During 1972, the growth in the volume of world trade accelerated to 9 percent and then surged by more than 12 percent in 1973. All of these rates are in real terms; the growth in dollar values was substantially higher, particularly in 1973, when the dollar devaluation and the worldwide surge of inflation helped lift the dollar value of world trade to a level 37 percent above a year earlier. (See *IMF* 1974 *Annual Report,* Table 5, p. 13.)

The expansion of world trade has tapered off since about mid-1973, along with the slowdown of real economic activity in the industrialized countries. More fundamentally, changes in the volume of world trade in the period 1971–73 appear to have maintained essentially the same relationship to movements in industrial production or real *GNP* as characterized the previous decade or even two, suggesting that the costs associated with the conduct of international trade have not risen significantly during the period of transition to limited exchange rate flexibility.[17] Similarly, although data on a worldwide basis are not yet available for 1973, the flow of long-term investment between the United States and other countries, at least, does not appear to have been adversely affected by exchange rate variability per se. The net flow of *U.S.* long-term investment abroad increased by more than 20 percent in 1973 over the 1972 level, and the net flow of foreign long-term investment into the United States increased by nearly 200 percent.

V. In Summary

In trying to sum up what we have learned about the process of adjustment to exchange rate change, we find that this period of transition leaves us with more new questions than new answers, that we must be tentative about everything and dogmatic about nothing. On the one hand, the experience of 1971–73 suggests to us that flexible exchange rates need not signal a retreat into autarky nor lead to beggar-my-neighbor policies of competitive depreciation and that, in the case of the United States at least, they do "work" as an instrument of balance of payments adjustment. That is, a depreciation, despite an initial perverse effect, does appear to lead to an improvement in the trade, current account, and basic balances, and despite substantial fluctuations, the exchange markets do appear to be dynamically stable. On the other hand, greater rate flexibility is not necessarily accompanied by a reduction in restrictions on international transactions; it gives considerably less than complete autonomy to domestic monetary policies; and rate changes appear to have a larger feedback onto the domestic price level than had been generally anticipated, at least in the case of the United States.

Flexible rates offer no perfect reconciliation between the desire for world market integration on the one hand and the demand for autonomy of national policies on the other; they are at best a second-best solution in a world where the efficiency advantages of integrated markets and a single international money must yield place to nations' insistence on individual sovereignty in the determination of domestic economic targets and policy instruments. Whereas we are uncertain about the impact of different exchange rate regimes on the course of world inflation, there can be no doubt about the relationship in the other direction: there can be no alternative to rate flexibility in the presence of high and widely divergent rates of inflation among nations. In fact, the very existence of such divergent rates of price increase indicates that countries regard the costs of maintaining a common rate of inflation, as required by a fixed-rate regime, as excessive in terms of their domestic economic goals. Furthermore, there is a widespread recognition of the fact that a system of fixed parities could not have weathered successfully the economic shocks of the past two years, and that the alternative to substantial fluctuations in foreign exchange markets would have been nonfunctioning markets for substantial periods. One might characterize the present exchange rate situation as Winston Churchill is said to have characterized democracy: "it is the worst possible system, except for all the others."

---

[17] Makin offers some detailed regression analysis in support of this proposition.

REFERENCES

J. R. Artus, "The Behavior of Manufactured Export Prices," paper presented at Conf. on Effects of Exchange Rate Adjustments, U.S. Treasury Dept., Washington, Apr. 4–5, 1974.

W. Branson, "The Trade Effects of the 1971 Currency Realignments," *Brookings Pap. Econ. Act.*, 1972:1, 15–58.

A. J. C. Britton, "The Dynamic Stability of the Foreign-Exchange Market," *Econ. J.*, Mar. 1970, *80*, 91–96.

P. B. Clark, "Effects of Recent Exchange Rate Changes on the U.S. Trade Account," paper presented at Conf. on Effects of Exchange Rate Adjustments, Washington 1974.

R. A. Debs, "Inflation and the Economic Outlook," *Fed. Res. Bank New York*, Apr. 1974, *56*, 83–87.

R. Dornbusch, "Devaluation, Money and Nontraded Goods," *Amer. Econ. Rev.*, Dec. 1973, *63*, 871–880.

O. Emminger, "Inflation and the International Monetary System," Tenth Lecture of the Per Jacobsson Foundation, Basle, June 16, 1973.

H. J. Gailliot, "Purchasing Power Parity as an Explanation of Long-Term Changes in Exchange Rates," *J. Money, Credit, Banking*, Aug. 1970, *2*, 348–357.

R. J. Gordon, "Interrelations Between Domestic and International Theories of Inflation," paper presented at the Wingspread II Conf. on the Political Economy of Monetary Reform, Racine, Wis., July 24–27, 1974.

G. Haberler, "Inflation as a Worldwide Phenomenon: An Overview," paper presented at the Conf. on the Phenomenon of Worldwide Inflation, Amer. Enterprise Instit., Washington, May 6–7, 1974, *4*.

P. Isard, "The Price Effects of Exchange Rate Changes: Some Evidence from Industry Data for the United States, Germany, and Japan, 1968–1973," paper presented at Conf. on Effects of Exchange Rate Adjustments, Washington 1974.

H. G. Johnson, "Problems of Stabilization Policy in an Integrated World Economy," in *Further Essays in Monetary Economics*, Cambridge, Mass., 1969, 338–352.

———, "Inflation and the Monetarist Controversy," *The Professor F. de Vries Lectures*, Amsterdam and London, 1972a.

———, "The Monetary Approach to Balance-of-Payments Theory," *J. of Fin. and Quant. Analysis*, Mar. 1972, 1555–72, 1972b.

———, "Secular Inflation and the International Monetary System," *J. Money, Credit, Banking*, Feb. 1973, 509–519.

H. B. Junz and R. R. Rhomberg, "Price Competitiveness in Export Trade Among Industrial Countries," *Amer. Econ. Rev. Proc.*, May 1973, *63*, 412–418.

W. Kasper, "The Effects of Exchange-Rate Change: Recent International Currency Experience," Canberra Aug. 1974, processed.

C. P. Kindleberger, "Lessons of Floating Exchange Rates," paper presented at the Conf. on Economic Policies, Carnegie Mellon and Rochester Universities, Pittsburgh, Nov. 15–16, 1974.

S. Y. Kwack, "The Effects of Foreign Inflation on Domestic Prices and the Relative Price Advantage of Exchange Rate Changes," paper presented at Conf. on Effects of Exchange Rate Adjustments, Washington 1974a.

———, "Simulation of Exchange Rate Effects with a U.S. Balance of Payment Model," paper presented at Conf. on Effects of Exchange Rate Adjustments, Washington 1974b.

A. B. Laffer, "The Phenomenon of Worldwide Inflation: A Study in International Market Integration," paper presented at Conf. on the Phenomenon of Worldwide Inflation, Washington 1974.

D. E. Logue and T. D. Willett, "The Effects of Exchange Rate Adjustment on International Investment," paper presented at Conf. on Effects of Exchange Rate Adjustments, Washington 1974.

S. P. Magee, "Currency Contracts, Pass-through, and Devaluation," *Brookings Pap. Econ. Act.*, 1973:1, 303–323.

J. H. Makin, "Eurocurrencies and the Theory of International Money," paper presented at Conf. on Eurocurrencies and National Financial Policies, Amer. Enterprise Instit. and U.S. Treasury Dept., Washington, Oct. 17–18, 1974.

R. I. McKinnon, "Floating Exchange Rates 1973–74: The Emperor's New Clothes," paper presented at Conf. on Economic Policies, Pittsburgh 1974.

W. Nordhaus and J. Shoven, "Inflation 1973: The Year of Infamy," *Challenge*, May–June 1974, 14–22.

L. Rosenberg, "The Impact of the Smithsonian and February 1973 Devaluations on Imports: A Case Study of Steel Imports," paper presented at Conf. on Effects of Exchange Rate Adjustments, Washington 1974.

W. S. Salant, "The Post-Devaluation Weakness of the Dollar," *Brookings Pap. Econ. Act.*, 1973:2, 481–496.

J. Salop, "A Note on the Monetary Approach to the Balance of Payments," paper presented at Conf. on Effects of Exchange Rate Adjustments, Washington 1974.

J. Williamson, "Exchange-Rate Flexibility and Reserve Use," *International Monetary Fund DM* 74/85, Aug. 29, 1974, processed.

Bank for International Settlements, *44th Annual Report,* Basle, June 1974.

Council of Economic Advisers, *Economic Report of the President,* Washington, Feb. 1974.

International Monetary Fund, *1974 Annual Report,* Washington 1974.

———, *IMF Survey,* various issues.

Morgan Guaranty Trust Co., *World Financial Markets,* various issues, and data sheets.

U.S. Congress (93–1), How Well Are Fluctuating Exchange Rates Working?, Hearings before the Subcommittee on Int. Econ. of the Joint Econ. Comm., June 20–27, 1973.

DISCUSSION

Rudiger Dornbusch, University of Chicago: Marina v. N. Whitman has provided us with a broad-based and quite eclectic review of our knowledge. Her conclusion that we must be tentative about everything and dogmatic about nothing is perhaps unduly pessimistic, and I shall therefore concentrate in my remarks on those issues that I feel we can quite confidently make strong statements about.

The first issue about which we can be quite dogmatic is that balance of payments questions, and more narrowly trade balance questions, are issues in the realm of macroeconomics and that consequently any approach that does not cast the problem in that setting is at best misleading and more likely incorrect. Nobody, I believe, is on record questioning this fundamental proposition about the adjustment mechanism, but still the literature and policy discussion abounds with instances where the implication of that assertion is evaded by approaches euphemistically referred to as "partial equilibrium," whatever that may mean in a macroeconomic context.

The recognition that payments problems are macroeconomic problems raises doubt about two propositions that have recently attracted interest. The first is the J-curve—the suggestion that a devaluation in the short run causes a worsening in the trade balance to be followed subsequently by an improvement. Already in the framework of a crude elasticity approach, there is no presumption that such an outcome is inevitable. More fundamentally, however, there would not seem to be a macroeconomic relationship that would predict an increase in nominal spending relative to income as a consequence of an exchange rate change, inevitably and independently of monetary and fiscal policy. It is true, however, that theories of the short-run consumption function might shed an interesting perspective on the problem and that accommodating monetary and/or fiscal policy would surely contribute toward establishing why we observe J-curves.

A second proposition that has to be questioned concerns the alleged permanent relative price advantage for exports that would be retained from a devaluation. This is indeed an extraordinary suggestion since it is argued that an exchange rate change in and of itself exerts permanent real effects. As it stands, the only support for such an effect is permanent money illusion. Alternatively, with the appropriate macroeconomic support, an exchange rate change can and will have long-run real effects. In that context, however, the roles are essentially reversed. A change in the balance between income and expenditure is essential and an exchange rate change is helpful in effecting the necessary relative price adjustments via an increase in the price of traded goods rather than a deflation-induced decline in the price of home goods.

A macroeconomic perspective suggests that changes in absorption that are implied by changes in the trade balance will have systematic effects on relative prices. As a first approximation, the relative price of domestic goods will decline in terms of all traded goods in the surplus country while in the deficit country, where absorption expands, the relative price of nontraded goods will increase in terms of traded goods. These relative price changes are required in order to translate changes in absorption, part of which falls on nontraded goods, into a change in the trade balance.

The preceding line of argument can be used to establish how an exchange rate change, accompanied by a validating aggregate policy, will affect the "competitive position" of a country in world markets. The relevant reasoning is not so much that domestic exports become cheaper abroad and imports become more expensive at home—the relevant distinction is one between traded and nontraded goods. The improvement in the competitive position of domestic producers of traded goods derives from the fact that the reduction in absorption frees factors of production in the home goods industries and thereby lowers real factor costs to the traded goods sector, making its expansion profitable. Conversely, in the deficit country the expansion in absorption and the nontraded goods sector causes tightness in factor markets and renders traded goods industries less profitable by raising real factor costs in terms of traded goods.

The enhanced competitive position of the home country does not only extend to established traded goods. In fact, the reduced domestic absorption causes some goods to move from the home goods sector to become exportables as their relative price decline makes them sufficiently competitive to overcome tariff and transport costs. Conversely, on the importable side the reduced absorption contributes to increasing the size of the sheltered sector. The distinction between internationally traded goods and home goods is therefore not excessively sharp but provides, nevertheless, the basis for systematic relative price effects of changes in absorption.

From discussion of the transfer problem it is known that there is no presumption about the terms of trade effects of a redistribution of income. By analogy the redistribution of absorption that is implied by a trade balance change does not affect in a systematic manner the relative prices of traded goods: the home country's exportables do not become, as a matter of presumption, cheap or expensive relative to importables.

None of the preceding remarks is in any manner at variance with the "monetary approach" to the balance of payments. That approach most assuredly does not require the "small" country assumption nor the assumption that particular and unusual relations obtain between prices of internationally traded commodities. All it does assume—as does indeed any approach I am familiar with, and in particular the elasticity approach and the Keynesian model—is that the prices for identical traded goods are equalized between countries with due allowance for exchange rates, tariffs, and transport costs. If they were not equalized, there would be unexploited profit opportunities to be eliminated by arbitrage. The misunderstanding would seem to result from the use, in some expositions, of one-commodity models that were chosen to drive home the point that exchange rate changes have nothing to do with terms of trade changes.

The discussion of flexible exchange rates in Whitman's paper raises several important questions. I entirely share her suggestion that stock considerations are important in the determination of exchange rates and should go a step further to say that exchange rates are determined in the market for asset stocks and not in the market for flows of foreign exchange. This point was noted by Joan Robinson when she remarked that the exchange value of a currency depends on the amount of it in existence with sufficient allowance made for changes in the demand for money. Such a view of the determination of exchange rates is obviously far removed from a conception of the exchange rate as balancing a current account with an independently specified capital account. A further implication of the stock market view of exchange rates is the substantial scope for expectations in the determination of the exchange rate. Anything that affects the demand or supply of money will as a first approximation be reflected in changes in the exchange rate.

The autonomy for monetary policy that is provided by flexible exchange rates has long been recognized as a principal attraction of that exchange rate regime, but has perhaps, too, been exaggerated. There is no question that in the long run flexible rates allow a country to choose its own rate of inflation independently of the rest of the world. Differences in nominal interest rates will reflect differences in anticipated inflation rates and the implied appreciation or depreciation of exchange rates. Beyond that autonomy there is very little. Flexible rates will not insulate a country against real disturbances originating abroad, nor will they insulate a country against foreign monetary disturbances since the latter will, at least in the short run, exert real effects. Capital mobility will ensure the transmission of monetary disturbances by keeping anticipated real interest rates in line internationally.

WILLIAM J. FELLNER, Member of the President's Council of Economic Advisers: I shall first ask and try to answer a very specific question connected with policy decisions. Thereafter I shall suggest a more general conclusion which bears on the central theme of Marina v.N. Whitman's valuable paper.

Say a country needs to get rid of a trade deficit and say that this needs to happen because an external source ceases to be available for further financing (for a further acquisition of claims or further subsidization). We shall assume that the need to balance the trade account does not arise with such suddenness as to exclude the bridging of lags during a transition period. Now, define *resistance* of the deficit country's economy to the equilibrating effort at full utilization as meaning that, at least temporarily, *underutilization* (unemployment) becomes a by-product of the trade-equilibrating process. In the analysis I suggest here, this could happen because it is in practice impossible to make money wage rates consistent with the requirements of full utilization. The first question I will pose is whether resistance in this sense is or is not apt to be greater if in the deficit country demand is reduced in terms of the domestic currency under a fixed rate system than if the exchange rate of the currency is allowed to become reduced in response to market forces and the demand for goods is kept unchanged in terms of the currency of the country. (The assumption concerning strictly unchanged nominal demand can and will be somewhat relaxed.)

Under a properly functioning fixed rate regime the adjustment process will involve either a reduction of after-tax money wage rates (I mean a reduction relative to trend) *or* it will involve underutilization. In either event nominal and real balances will presumably be reduced. On the other hand, under a properly functioning regime in which the exchange rate of the deficit country is allowed to decline in response to market forces, the process is compatible with unchanging after-tax money wage rates and other after-tax money incomes and with a one-time increase in consumer prices expressed in terms of the domestic currency. Given the unchanging after-tax money incomes and the higher prices, trade will become balanced here too, with smaller imports and higher exports and presumably with smaller real balances (unchanging nominal balances). The process could even be described as involving a one-time increase in money wage rates that falls short of a one-time increase in the price level. From this it, of course, does not follow that in the case of rate adjustment there will be no resistance in the sense in which the term is here used, since *real* wage rates *will* have declined in either case. If for the time being the real decline is accepted only in a state of underutilization then the outcome will reflect resistance (in our sense) even in the case of rate adjustment. The first question that needs to be considered is, however, whether a difference in the degree of resistance can be expected when we compare these two ways of eliminating the trade deficit and, if so, whether the difference is to be attributed to mere "money illusion."

As I see it, a difference *is* to be expected quite aside from any money illusion that may exist. The reason for this is that, in countries in which collective bargaining plays an important role, a concern with *relative* real wage positions nevertheless gives the money wage bargain a special status. No union management suddenly called upon to accept a burden in terms of money wages knows what is about to happen to the relative real wage positions that have gradually evolved in the economy. This is true especially because a good many negotiated money wage rates are not even capable of being changed in the short run. On the other hand, it is true in an acceptable first approximation that, if the real burden is imposed in the form of higher consumer prices, real wage rates are reduced in the same proportion across the board.

This same distribution of the burden could "in principle" be achieved even under the fixed rate regime, but only if in the deficit country the deflationary effect were achieved by taxing all wage earners (and presumably other income recipients) in the identical proportion and by "destroying" the money so collected. Such changes in the tax structure could clearly not be achieved with the required promptness, all the less because only gradually could it be discovered what the size is of the needed taxation. The reserve-flow mechanism is rightly represented as working not through such tax changes but through pretax money income adjustments, and these do pose the question of how easily money wage rates give way piecemeal with unpredictable consequences for relative real wage positions. Thus it is reasonable to expect a higher degree of resistance at full utilization in the event of the fixed rate than in that of the flexible rate adjustment process.

The reasoning so far has assumed that neither the flexible rate system nor a system aiming for fixed rates is misused by governments with inflationary proclivities. However, postwar history has demonstrated that governments have a strong inclination to generate inflationary processes which provide a temporary stimulus to economic activity. Such policies provide a stimulus as long as the actual rate of inflation can be kept higher than the expected inflation rate with the result that the available real incomes are overestimated by the public. This has been happening even though the price that needs to be paid for such policies subsequently is uncomfortably high. The flexible rate system can be misused by governments with such inflationary proclivities and systems based on other principles also can be. How?

In a flexible rate system a deficit country can engage in inflationary policies by which money wage rates are allowed to rise progressively at high utilization levels in circumstances where an underestimate of the inflation rate by the public—hence an overestimate of real wage rates—creates the belief that an inevitable burden is being avoided. The one-time price increase that would take place under a properly conducted flexible rate adjustment process degenerates into an accelerating inflationary process to which at a later stage it will be necessary to put an end at a considerable cost in terms of the rate of resource utilization. Meanwhile rate adjustments will presumably repeat themselves. A flexible rate system misused by inflationary practices is clearly inferior to a fixed rate system that is not misused by governments with strong inflationary tendencies.

Yet there exist methods by which systems based on the fixed rate principle can also be so misused, and if misused, these systems are also apt to lead to very bad results.

One such method of misuse is that of endowing a fixed rate system very richly with reserves and of letting the adjustment processes develop exclusively by means of inflation in the surplus countries. This imposes more inflation on the surplus countries than the amount which their own political processes would make inevitable. And given sufficiently ample reserves, there is no presumption that the trend in the deficit countries would become less inflationary than the initial political pressures in those countries would have required. This is one method of misusing a fixed rate system.

The other method by which systems based on the fixed rate principle can be misused is that of introducing escape clauses that enable countries to administer abrupt exchange rate modifications in circumstances in which flexible rates would have moved gradually. When misused by either of these methods, or by a combination of the two, such a system yields no better results—probably yields worse results—than a misused flexible rate system. The past provides many illustrations for the misuse of both systems.

It is essential to ask oneself the question what a government is capable of achieving if it has learned from past experience and has developed a firm intention not to misuse the currency system, though it may prove incapable of standing up against political pressures if these become too acute. I suggest that in the contemporary environment the pressures against which a competent government trying to play the game according to the rules will have to stand up are greater under a system aiming for fixed rates than under a system leaving room for rate flexibility in response to market forces. Standing firm against pressures to generate a self-defeating inflationary process, difficult as it may be, should nevertheless be much more manageable than generating a deflationary process which hits the individual elements of the wage structure piecemeal, and partly with a considerable lag, as wage negotiations come up for renewal.

In an uneasy world we have arrived at an uneasy case for flexibility. But the case is a lot less uneasy than is that for a system that would try to establish fixed rates in the present circumstances.

ALEXANDRE KAFKA, International Monetary Fund and University of Virginia: As the only underdeveloped member of this panel, I should like first to say a few words about how different exchange rate regimes may affect many of the less developed countries (*LDC*). I shall then make some general observations on our subject.

The *LDC*s have been concerned because the progressive abandonment of the par value system, as Marina v. N. Whitman shows, has meant not only medium-term realignments of average exchange rates among industrialized countries but also large short-term fluctuations. A par value system (as distinct from a fixed rate system) could not have avoided the medium-term realignments, but could have avoided the short-term fluctuations.

What are the problems which the floating of the currencies of industrialized countries creates for some *LDC*s? Those among them who do not peg to another country's currency may face problems of learning and institutional adaptation, involving at least transitory additional social costs, when major currencies fluctuate to a larger extent than under the par value system. First, an exchange rate policy may have to be instituted, i.e., decisions may have to be taken whether to peg to a single currency, to a basket of currencies, allow the currency to float freely (whatever that may mean), etc. Second, forward markets may have to be created where they previously did not exist. This can be avoided by pegging to a single currency; and such pegging may in any case be unavoidable for reasons of economic size and tradition of the pegging country. But pegging to a single floating currency may involve disagreeable political decisions as well as balance of payments and internal stabilization problems additional to those faced under a par value regime.

*LDC*s also fear that they may be confronted by a problem of reserve supply if major currencies continue to float. John Williamson has suggested that countries pegging to a floating currency (which is the case of most *LDC*s) are likely to experience increased reserve use and reserve needs. The reason is the likely increase in the variability of their receipts and payments, if the currency to which they peg fluctuates in terms of third currencies and if their trade is partly with third countries. If, by contrast, increased exchange rate flexibility among major countries reduces or, at least, does not similarly increase their reserve needs, the propensity to create reserves for the world as a whole will become inadequate from the point of view of *LDC*s pegging to a floating currency.

As against these inconveniences, *LDC*s would very much agree with Whitman that the recent type of floating has not clearly led to reduced restrictions—though admittedly nobody knows what would have happened in the absence of floating.

I agree with almost—but not quite—everything else that Whitman says in her truly excellent paper about the subject of our panel. I have three areas of doubt.

First, her description of the present system as a managed float (with relatively little intervention) is certainly correct. Perhaps, however, the management is, so far, even a little less disorderly than she suggests. Intervention takes place subject to consultation and is not at cross purposes. Even other policies are applied to some extent in accordance with internationally agreed principles. Thus, international borrowing is in accord with the principle that oil deficits as a whole must be financed at this time since to attempt to adjust them would be futile. We have also a set of internationally agreed guidelines for floating designed to avoid competitive exchange rate changes and, in addition, to permit the International Monetary Fund (*IMF*) to encourage (but not to oblige) countries to keep rates somewhere within a sensible range in relation to each other if there should be a tendency which might move them away from such a range. So far, these guidelines have at least not been obviously flouted by any country. Admittedly, however, the temptation to do so has been smaller than it is likely to be in the near future.

That we should have a managed float is hardly surprising. Since modern governments bother about prices, they can hardly be expected not to bother about exchange rates and not to try to control them. One aspect of this concern refers to the effect of the exchange rate on the structure of the balance of payments: if adjustment is possible, a deteriorating current account may be bothersome even if overall equilibrium is preserved due to capital inflows with an appreciating exchange rate. From the international point of view, and contrary to the conventional view, "free" floating is even more difficult than managed floating to police as an international system. Anything a government does—not just exchange market intervention—is likely to affect the exchange rate. And one can never be sure whether a policy was designed to alter the exchange rate or not, so that one can never be sure whether the adoption of any policy contravened an agreement not to manage a float. Under managed floating, there can at least be an agreement to encourage or even oblige a country to keep its exchange rate within a sensible range.

Second, the system has not avoided large short-term fluctuations of rates. But—and even more important—one may also doubt its ability to bring balances of payments closer to equilibrium. Whitman shows that it has done so for most balance of payments concepts in the case of the United States (except for the oil impact) but, as C. P. Kindleberger reminds us in a recent lecture, emphatically not in the case of Germany (despite the oil impact). Did effective exchange rate movements not produce the same effects in Germany as in the United States? Or did the deutsche mark effective rate, for some reason, fail to move adequately (considering not only divergent price movements but also, e.g., divergencies in delivery terms)? Was such a failure due to slight but misplaced intervention outside the mini-snake, or to the persistence of some controls affecting capital movements, or to the need to give speculators time to learn how to speculate under floating rates? One must hope it was not the latter, as recent rate movements do not suggest that much learning is taking place. Or must one simply conclude that major divergencies between countries in demand management policies and cost-push, such as existed between Germany and her major trading partners, are apt to mean that fluctuating exchange rates do not work well? But par values also work well under ideal conditions.

My third doubt concerns the relationship between worldwide inflation and the exchange regime. Floating rates, including rates subject to some management, would, according to traditional theory, be expected to bottle up inflation (or deflation) in the country where it originates (and give other countries greater policy autonomy). Of course, worldwide inflation does not imply that inflation has spread, despite floating; domestic inflationary pressures may have just happened to be simultaneously present in countries whose currencies were floating. In fact, however, floating rates may actually contribute to the spread of inflation, a point made first, I believe, by Marcus Fleming and which reinforces Whitman's own skepticism. Theory teaches us only that floating rates would isolate economies in the absence of capital flows. If, however, capital is sufficiently mobile, a boom and inflation in one country may attract it in such large quantities that that country's exchange rate appreciates instead of depreciating. Thereby it spreads the boom and inflation to other countries even more than would be the case under fixed rates and gives other countries even less policy autonomy.

Where does this leave us? Certainly not necessarily with a preference for a par value system at this time, but with a preference for a system which replaces par values by an obligation to maintain exchange rates not within fixed margins but within zones with soft margins. In other words, a more specific obligation than that involved in *IMF* guidelines might well be preferable: vouchsafe more stability than the present system and yet handle at least as well exchange rate changes required in present unsettled conditions. The more specific obligations mentioned may appear more necessary as temptations to flout the *IMF* guidelines increase.

Kindleberger in a recent lecture spelled out the difficulties one meets in making up one's mind about how well any given system has worked, let alone about how well it is likely to work in the future. But even apart from his warnings, on the grounds given earlier, it seems to be difficult to agree with Whitman's conclusion that the system which has emerged "is the worst possible system except for all others." One will have to be even more undogmatic than she is. There is no proof that any other system would have been better, but there is certainly none that all others would have been worse.

Paul A. Volcker, Princeton University: Marina v. N. Whitman seems to me properly modest in summing up for the profession what it has learned so far from the recent experience with flexible exchange rates. Indeed, a practitioner in the area might be excused for noting a sharp contrast between the tone of authority with which most economists pronounced policy judgments about flexible exchange rates on the basis of theoretical constructs before 1971 and the tone of humility that emerges from attempts to assess actual experience with such a regime since that time.

To be sure, Whitman reassures us that, "in the case of the United States at least," the evidence is that movements of exchange rates do have the anticipated effects on the direction of the current account adjustment. Exchange markets therefore appear "dynamically stable," at least in the limited sense that exchange rates, left to themselves, will not oscillate between zero and infinity. But we are still left pretty much in the dark about the relevant elasticities and lags involved in exchange rate changes.

In the midst of this "scientific" uncertainty, the disconcertingly large fluctuations in market exchange rates over the past two years are perhaps not surprising.

Indeed, the now properly fashionable technique of calculating trade-weighted "effective exchange rates" obscures, as Whitman implies, the degree of instability we have seen in important exchange markets. Movements in "effective rates," particularly for the dollar, are dampened by the practices of some countries in formally or informally pegging their currency to the *U.S.* dollar, or by the relative stability of rates between some closely linked economies where historical "parity" relationships seem normal. Meanwhile, exchange rate relationships between the most important world traders—the United States, Germany, and Japan, each of which compete in world markets over a broad range of manufactured goods—have gyrated through a range that, on the face of it, bears little resemblance to changes in domestic price and cost structures.

For instance, deutsche marks today buy 65 percent more dollars than they did in 1969; and at one time in 1973 about 75 percent more. Yet, in between, they temporarily lost over 20 percent of their peak dollar value. Perhaps more understandably, the yen appreciated at one point by 42 percent vis-à-vis the dollar, but then fell back by over 12 percent in the space of a few weeks. We can find more extreme examples among the cross rates for such important and close trading countries as Italy, the United Kingdom, Switzerland and the Netherlands.

If economists are highly uncertain about the size and timing of the impact of exchange rate changes, why should they (of all people) expect stabilizing speculators—some never very clearly defined group with access to large resources—to be quick to put not just their professional judgment but real money on the line? At least on a bandwagon you have the psychological comforts of good company and loud enthusiasm in return for the price of admission.

Of course, we must not forget we are in a transitional and learning period. Disequilibria had accumulated over years, and the world economy has been buffeted by unprecedented disturbances from oil, food, and inflation. In time we should be able to look to the economist for more reliable answers to the open questions—and to businesses and banks for more efficient speculative stabilization. But I do not know how much time we have. Until recently we have had a high level of mutual prosperity and common concerns with shortages and inflation. These provided powerful incentives to maintain open import markets. Businesses, pressed to deliver and obtain materials, have worried more about availability than price. As we turn to preoccupation with recession, unemployment, and payment for oil bills, new doubts and questions seem likely to arise about how well floating exchange rates reconcile the desire to maintain a high degree of international integration, ease adjustment problems, and preserve autonomy for domestic policies.

Concern over the present functioning of the market is not enough. What is the alternative? The present state of economic knowledge seems to support the instinctive reaction on nearly all sides to rule out an early return to a fixed rate regime. Whitman has demonstrated how little we would have to go on in establishing new fixed rates, and in changing them at discrete intervals, especially in a world where there is no *n*th country willing and able to play a relative passive role, absorbing any residual deficit that might result from letting other countries "play it safe" by resolving their own doubts on the side of surpluses. The credibility of new fixed rates would require a commitment of official financial resources and a willingness and ability to coordinate internal policies to a degree that simply does not now exist.

But conversely, the elasticity doubts and the possibly long lags associated with adjustment to exchange rate changes—to say nothing of new doubts raised about the internal price repercussions of exchange rate fluctuations—reinforce concern that large overshoots may be characteristic of more or less free floating and bring in its wake large and unnecessary dislocations in both domestic and international markets. Because our foreign trade sector is so small (although growing) relative to the total economy, and because of concern that official intervention can be overwhelmed in the context of large and open capital markets, the temptation has perhaps been strongest in this country to live with these uncertainties and distortions as the price of domestic autonomy, meanwhile awaiting more settled conditions and the stronger appearance of the benign speculator. But Whitman makes clear we are in the midst of a new debate over how much autonomy even the United States has gained.

We have learned, for instance, that in conditions of strong demand and price pressures, with important export- or import-competing industries already experiencing capacity and shortage problems, a depreciation may result in more price pressure and less balance of payments adjustment than bargained for.

In the end balance of payments adjustment (to the extent it cannot be achieved through capital accounts alone) requires a shift of resources, and that shift has domestic as well as international consequences. In a given economic and political structure (say the United States) it may sometimes be easier to accomplish that shift with less pain (real or observed) if the shift initiates in the relatively obscure manner of changes in import and export prices. At other times, or in other countries, the relative advantages may be different—and a monetary *system* must serve the needs of more than one country.

Whitman questions particularly whether the vaunted independence for monetary policy in a regime of floating rates has not turned out to be a good deal less than promised. No doubt it will be a long time before we have the final word on this subject. But observation of recent events suggests we have a slower and more uncertain real adjustment process, a much larger supply of interest-sensitive (and bandwagon) capital relative to changes in current accounts, and less courageous benign speculators than most theorizing about independence for monetary policy presumed.

Whitman has put her finger on the weakness of expectations concerning future exchange rates as explaining the absence of stabilizing capital flows and the instability of observed exchange rates. A little later, she argues uncertainty concerning future rates is crucial to independent monetary policies. If she is right, we can't have it both ways—floating cannot at the same time give us reasonable stability and autonomy. Some clarification is needed, but in the end, I cannot square the circle—we will have to compromise.

Take the case where future expectations are weak or absent. Disturbance of an equilibrium by a change in relative interest rates, providing an opportunity for a known gain in interest returns, will presumably stimulate some uncovered capital movements and thus a change in present exchange rates. The initial exchange rate change is as likely under these conditions to attract followers as counterspeculation. If the interest differential persists, so should the pressures on exchange rates—and the market may well anticipate the changes in abrupt movements.

The process will presumably be self-limiting, but only at considerable cost. Substantial exchange rate changes would be necessary to induce strong expectations that the existing rates are too high or too low, ending incentives to net capital flows. Eventually, large changes in current accounts would develop to the point that real rates of return on capital are affected, a new equilibrium in interest rates forced, and the process reversed. Obviously, this kind of corn-hog cycle in exchange rates would be cumbersome and the costs in terms of international integration high. In specific circumstances, a country easing or tightening monetary policy may willingly accept or welcome the consequences for the current account as reinforcing the objectives of monetary policy. But carrying that process very far invites "beggar-my-neighbor" (or "inflate-thy-neighbor") descriptions.

Now take the case of strongly held future expectations. In this case a "collar" is placed on the deviation of the present exchange rate from the future expected rate, roughly measured by the relevant interest rate differentials. At first blush, this seems to be ideal, for domestic monetary autonomy is achieved with limited exchange rate repercussions. The trouble seems to me that, if the future expectations are not borne out by experience, they eventually will be dropped. And those expectations are not likely to be borne out if monetary policies, among other factors, remain divergent. It is in that sense that Whitman seems to me correct that certainty about future exchange rates is inconsistent with autonomy in domestic policy—it promises more than can be delivered.

I am forced to the mundane conclusion that, in an integrated world, domestic policy autonomy is a will-o'-the-wisp, whatever the exchange rate regime. What does of course differ is the manner in which a domestic policy initiative or other disturbance affects different sectors of the home economy and is transmitted abroad—particularly the degree to which adjustment is forced initially on the foreign trade sector. Different countries will perceive their interests differently in this respect, so compromise is called for. At the same time, nobody has an interest in the repetitive undershoots and overshoots inherent in unformed expectations. Both considerations suggest a strong common interest among the relevant authorities in providing more leadership toward solidifying future exchange rate expectations than has been apparent.

In general terms, my proposition is that floating needs management; because we are starting in such disturbed circumstances, the management needs to be substantial; and to be credible and effective, the management needs to extend beyond direct or indirect exchange market intervention into all the old ground of coordination of monetary and other policies. In time, as our domestic economies become more stable, and as we and the markets learn more about what exchange rates are consistent with balance of payments equilibrium, we might anticipate the need for management, certainly in the sense of intervention, to progressively lessen. At that happy stage, the operational distinction between managed floating (soft bands) and readily adjustable pegs (hard bands) could more or less wither away.

I do not suggest this is analytically, operationally, or politically an easy approach. It does not have the simplicity of the free floating or the fixed options. If governments and central banks are to step up to take the risks that stabilizing speculators now refuse, and if they operate with some intelligence and understanding, they should be rewarded over time by financial gains from performing an important social function. But the risk of financial loss (to say nothing of loss of face) is apparent. However much economists may be inclined to dismiss the matter as bookkeeping, this matter is a large deterrent to action—perhaps more in the United States than elsewhere.

Assessment and sharing of these risks lead directly into the broader question of the division of responsibility (and judgment) for intervention, and for reasonable coordination of monetary policies, between nations on the opposite sides of the same exchange rate. The convenience of an *n*th country, passive in exchange rates but powerful in monetary policy, will likely disappear as part of the process because it no longer fits the economic power relationships or the political sensitivities. Fortunately, the really key actors are relatively few, and they can approach the problem with a good deal of informality. But, as the process becomes multilateral, the whole range of issues involved in formal international monetary reform arises.

My responsibility today is for a brief comment on what we have learned from recent experience, not an essay on monetary reform. So I will conclude simply by saying that I see nothing in what has happened that detracts from the need and urgency to complete the reform effort—or that makes the job seem any simpler than before.

25 YEARS AFTER THE REDISCOVERY OF MONEY: WHAT HAVE WE LEARNED?

Recent Developments in Monetary Theory

By STANLEY FISCHER*

The revolution in the significance attached to the role of monetary factors in the determination of the price level and rate of economic activity in the last twenty-five years is empirical, not theoretical. Comparison of Franco Modigliani's 1944 exposition of the Keynesian system with Milton Friedman's 1970 framework for monetary analysis reveals little change in the basic short-run macroeconomic model. Nonetheless, monetary theory today is very different from that of twenty-five years ago; the major change lies in the shift from comparative static analysis to explicitly dynamic analysis of the impact of exogenous variables on prices, interest rates, and, where variable, employment.

Because Harry Johnson's excellent 1962 survey comprehensively covers earlier developments, I will concentrate in this paper on those branches of monetary theory which have undergone most of their development over the last decade. In Sections I–III, respectively, I discuss work on the new microfoundations of money; disequilibrium theory; and money, price dynamics, and growth. The rational expectations hypothesis, and its implications for price dynamics, is taken up in Section IV, with conclusions presented in Section V.

There are also some noteworthy omissions. The most significant from the viewpoint of macroeconomics is the new microeconomics of unemployment and the Phillips curve which is not discussed since money is not directly involved in most of the models. Space limitations force me to define developments in the theory of policy and applications of control theory to macroeconomics as empirical; further, given my emphasis on the novel, I do not discuss some important extensions of earlier work on portfolio theory, the demand for and supply of money, and the term structure of interest rates.[1]

I. The New Micro Foundations of Money

Monetary analysis typically starts from an informal and usually picturesque explanation of the use of money, involving the double coincidence of wants and the four functions of money. A demand for money is then postulated and money is treated as one of the assets in a macro model, distinguished from other assets in one or more ways: since money is the medium of exchange demanded for its command over goods and services, the demand is a demand for real balances;

---

* Associate professor of economics, Massachusetts Institute of Technology. I am grateful to Jacob Frenkel, Milton Friedman, Pentti Kouri, Huston McCulloch, Franco Modigliani, Mary Kay Plantes, William Poole, Robert Solow, Ross Starr, Lars Svensson, and Hal Varian for helpful discussions and comments.

[1] Space limitations also require the number of references to be limited. A longer version of this paper with a more complete bibliography is available from the author on request.

since it is a medium of exchange, the real income elasticity of demand is positive; its interest return may be fixed exogenously; its nominal quantity, if it is a fiat money, can be exogenously determined.

The criticism "that the foundations of monetary theory have not yet been laid" turns primarily on the argument that there is no role for money in macro models, which are all small Walrasian general equilibrium models.[2] Essentially, it is argued that there is no role for money in intertemporal general equilibrium models since all trades in such models can be arranged in the first period the auctioneer opens the market; alternatively, within each trading period, there is no obvious need for a medium of exchange since all trade takes place with the market at equilibrium prices. Even if it is assumed that trade is between individuals rather than with the market, the theory offers no reason why one good rather than another should be used to balance trade between two individuals whose needs do not doubly coincide—nor does it offer a good reason why trade should balance between individuals. While these are compelling criticisms of the general equilibrium model as a description of modern economies, their practical import applied to the standard macro models is less clear, since several features of the demand for money in such models do reflect the medium of exchange function of money.

Recent literature on the foundations of monetary theory has attempted to make precise the notion of the double coincidence of wants; to present criteria for the selection of the medium or media of exchange (also, under the guise of "portability," "divisibility," etc., a concern of the early monetary theorists); and, fundamentally, to explain how it is that the use of money enables an economy to attain equilibria which are not possible without money. Three approaches can be distinguished. The first takes equilibrium prices for a single period as given and investigates the possibility and number of exchanges required to attain the implied equilibrium allocation of goods under various constraints on the exchanges that can be made when all trading is strictly bilateral. The second examines single period exchanges in an economy in which transactions are costly and determines market equilibrium and the nature of exchanges for that economy. The third uses a multiperiod general equilibrium framework in which exchange within periods is with the market, but in which money can be useful in making intertemporal transfers. Transactions costs and/or informational considerations are involved in all the analyses. Many of these analyses can be viewed as developments of work by Sir John Hicks, attempting to relate traditional explanations of the use of money to Keynes's three motives for holding money.

In most of the contributions in this area, money makes possible the attainment of equilibria which are otherwise unattainable by permitting the unbalancing of budgets in goods within a period or transaction. The reasons for balance at each transaction are elucidated by Joseph Ostroy and have to do with restrictions on the information which an individual is likely to have about other traders' solvency and the saleability of goods taken in exchange. Those models, such as Hahn's, which concentrate on intertemporal market equilibrium, do so because it is argued that the sequential nature of exchange is important, but then implicitly allow simultaneous trade within a period.[3] Indeed, they provide more reason for the use of bonds than for the use of a medium of exchange.[4]

---

[2] See Frank Hahn (1973a, p. 233). This article is a useful survey of one branch of the modern literature.

[3] See Hahn 1973b. An influential paper by Roy Radner emphasizes the sequential nature of transactions and the importance of the unfolding of new information over time.

There are few formal attempts to explain the emergence of money from barter systems; Mordecai Kurz argues that this social choice is fundamentally not analyzable by conventional economists' methods. Analysis of the structure of transaction costs leading to the adoption of some commodity as money, as in Jurg Niehans, is at least likely to restrict the choice of media of exchange, and work by Robert Jones provides an interesting start on the dynamics of the emergence of a medium of exchange. The literature mentioned so far has assumed nonincreasing returns to scale in transactions, though there are attempts to study money in nonconvex transactions technologies. The relationship between the existence of organized markets and the use of a medium of exchange is as yet unclear. Nor has uncertainty been given a prominent role in any of these models, except in Karl Brunner and Allan Meltzer.

This work is obviously both difficult and only a beginning. It is not clear where, if anywhere, it will lead. It will no doubt provide more convincing and carefully worked out explanations for the use of a medium of exchange than we now have, but it appears that those explanations will not be fundamentally different from the traditional verbal explanations, and that they will not have any major consequences for the way in which macromodels are built.

II. Disequilibrium Theory

The interpretation of the Keynesian system as an equilibrium system, except in the labor market, had become widely entrenched in the textbooks by the 1960's, in the shape of the familiar IS-LM apparatus. The apparatus was strongly criticized as essentially non-Keynesian by Robert Clower and Axel Leijonhufvud. I shall avoid the exegetical issue since the Clower-Leijonhufvud criticisms of the income-expenditure model, which have been followed by extensive work interpreting the Keynesian system as a disequilibrium system, are of independent interest.

Leijonhufvud's fundamental criticism is that income-expenditure theory suffers from auctioneer illusion. The informational problems the Walrasian auctioneer solves in obtaining an equilibrium price set are not, Leijonhufvud says, costlessly or immediately solved in the real world; they produce disequilibrium, trading at false prices, and unemployment—and, further, informational problems are complicated by the use of money. In particular, for reasons explored in the Edmund Phelps volume, quantities adjust more rapidly in response to disturbances than do prices and trade accordingly takes place at disequilibrium prices.[5] Disequilibrium trading produces deviation-amplifying (multiplier) effects on quantities which are absent from the typical Walrasian model. The work to be discussed below makes the basic Leijonhufvud position more precise.

The distinction between notional and effective demands is critical in disequilibrium analysis. Notional demands and supplies are calculated by individuals and firms on the assumption that they can buy and sell as much as they want of any good at prevailing prices. This assumption can be true only in equilibrium. In disequilibrium there are constraints on the amounts of some goods an individual can buy or sell. Effective demands are calculated conditional on the quantity constraints, or any other nonprice constraints, the individual perceives. For instance, if there is an excess supply of labor, the effective demand for goods by an individual who cannot sell as much labor as he wants is less than the notional demand for goods at prevailing prices. Further, his demand for goods in that situation will be a function of the amount of labor he can sell. This is the essence of Clower's dual decision hypothesis in which one set of demands is expressed if markets clear and another set if individuals are constrained. Clower applied his analysis to the consumption function, thus justifying again the Keynesian consumption function with income rather than the real wage as an argument.[6]

---

[4] It may well be, though, that only trivial transaction costs account for the failure of bonds to be used as medium of exchange. See Walter P. Heller.

[5] The Phelps volume (1970) was published after Leijonhufvud's book, but the article by Armen Alchian in the former is repeatedly referred to in the latter.

The most useful definition of effective demands is not yet settled. For instance, Herschel Grossman assumes that individuals express demands in each market which are based only on the quantity constraints perceived in other markets. Thus the effective supply of labor when the individual is constrained only in the labor market is the notional supply. This is a convenient but not compelling hypothesis. It also implies that effective demands do not necessarily obey Walras' law.

An analysis of the behavior of firms in demanding labor when they are unable to sell all the goods they want to at the prevailing price level had earlier been provided by Don Patinkin (Patinkin, ch. 13). In that situation firms' demand for labor is a function of the level of demand rather than the real wage. The Clower and Patinkin analyses are combined by Robert Barro and Grossman in a simple model with one perishable consumption good and money; households supply labor, purchase goods, and can attempt to save by adding to real balances. Firms hire labor to produce goods and both firms and households take prices as given.

For any specified wage and price levels, except at the Walrasian equilibrium, firms and/or households will be constrained in some markets. In particular, on the assumption that quantities are determined by the minimum of effective supply or demand in each market, there exist wage and price levels at which there are excess effective supplies of both labor and goods.[7] In such situations, the Keynesian cross becomes the model of output determination, and any change in demand (caused, say, by an increase in the nominal money supply) has the full multiplier effect on output—provided the change does not move the system into a different configuration of excess effective demands. Significantly, for this Keynesian case to prevail, the real wage may even be below the Walrasian equilibrium level if the absolute price level is so high as to reduce the effective demand for goods below the equilibrium level, through the real balance effect.

What does money have to do with all this? Leijonhufvud argues that the fact that the offer to work takes the form of a demand for additions to money balances rather than goods is crucial in generating the disequilibrium states. In a one-factor model in which the only market is one in which labor trades for a single good, it is, of course, possible to have unemployment if the real wage is too high. It is also clear that unemployment can arise at disequilibrium prices—and probably even at a real wage below its Walrasian equilibrium level—in models with many goods but no money. The role of money in disequilibrium trading is studied by Jean-Pascal Benassy. In any event, it is clear that the argument that money is crucial in generating disequilibrium cannot be studied in a model with exogenous prices.

---

[6] The classical life-cycle and permanent-income theories do not explicitly include the labor/leisure choice and thus treat income as predetermined.

[7] The fact that households can plan to save by adding to real balances is important in this regard.

Expectations are obviously important in disequilibrium analysis. To the extent that current unemployment is regarded as transitory, effective demand will be little reduced by constraints in the labor market as individuals draw down assets, if they have them, to maintain consumption; these expectational effects are made explicit in Barro and Grossman's forthcoming book in which the consumer's choice problem is set in a life-cycle context. The role of capital market imperfections and the distribution of unemployment relative to asset holdings of the totally unemployed in determining the size of the multiplier are studied by John Flemming.

The Walrasian assumption that prices move instantly to clear markets is replaced in disequilibrium analysis by the assumption that quantities can move instantaneously while prices remain fixed. Thus a discrete change in the stock of money or expectations, say, results in instantaneous movements in output in these models. The outstanding problem in macroeconomic disequilibrium theory is the working out of joint price and quantity determination.[8] The large search-theory oriented literature on price determination by firms which involves uncertainty—not so far embedded in the macrodisequilibrium analyses—will undoubtedly play a role in this synthesis. While disequilibrium analysis has succeeded in presenting models in which many Keynesian notions—particularly that of effective demand—are clarified, it remains to be seen whether the same analytical structure will prove useful when applied to situations in which price determination is endogenous.

III. Money, Price Dynamics and Growth

Recent emphasis on the dynamics of the adjustment to changes in the stock of money and its growth rate can be regarded as a return to themes of David Hume and Irving Fisher. Since much of the recent work has been set in a growth framework, I shall begin by outlining the framework and the uses to which it has been put. The major questions discussed in the money and growth literature are the "superneutrality" of money—in particular, the relationship between the steady-state rate of inflation and capital intensity—and the stability and more detailed dynamic properties of the processes of inflation and capital accumulation.[9]

In the simplest neoclassical growth model, as presented by Miguel Sidrauski, in which there are only one good and two assets, money and capital, equilibrium capital intensity is greater the higher the steady-state rate of inflation (Sidrauski 1967b). Money is injected into the economy through government transfer payments and the basic result occurs because a fixed savings rate is applied to disposable income: the injections of money have to be saved and displace some capital from portfolios. This result is due initially to Tobin (1965).

The basic Tobin-Sidrauski result can be reversed in descriptive models in a number of ways. First, if income in the monetary economy includes the imputed services of real balances, savings may fall sufficiently as the rate of inflation rises to offset the concurrent portfolio shift towards capital which otherwise increases capital intensity (Johnson 1969). Second, the result can be reversed by including real balances in the production function—as shorthand for the fact that the use of money in transactions frees real resources for the production of goods.[10] Third, the result depends on the assumption that money is of the outside variety.

---

[8] Barro and Grossman in their book append the "law of supply and demand" to their static model to discuss the dynamics and stability of their system which may not converge to the Walrasian equilibrium. Assumptions about the demands expressed in markets which do not clear—discussed in the above text—are obviously important in this regard. See Katsuhito Iwai for an analysis of joint price and quantity determination.

[9] Money is superneutral when steady-state real variables are independent of the growth rate of money.

The dynamics of inflation in monetary growth models depends heavily on the assumed process of expectations formation and the implicit or explicit reasons that prices change in the models. In neoclassical models, prices change to maintain portfolio equilibrium—to equate the demand for nominal balances, which may itself be represented by some distributed lag scheme, to the supply of nominal balances. In Keynes-Wicksell and other models, prices may change because they are expected to and/or because there is excess demand in the goods markets. Similarly, a wide variety of plausible expectations assumptions has been used. The way in which prices, real and nominal interest rates, and (where variable) the unemployment rate adjust to monetary changes depends on the details of the price-adjustment and expectations assumptions; it appears that almost any pattern of adjustment can be obtained by a suitable and not implausible choice of expectations and price-adjustment equations. Thus in this area, which is crucial from the policy viewpoint, existing theory seems to offer too much—the need is to restrict the class of acceptable results. Fortunately, certain restrictions are implied by the stylized facts emerging from the extensive empirical work in this area.

The indeterminacy of the basic results of the descriptive growth models, as well as the view that models should be based on optimizing behavior, leads naturally to attempts to restrict behavioral functions by appealing to models of individual optimization. In the initial such optimizing model, in which money enters the utility function and maximizing is over an infinite horizon, the steady-state capital stock is invariant to the rate of inflation (Sidrauski 1967a). This strong result is not, however, especially robust: treating money as a factor of production, or making the labor/leisure choice endogenous changes it.

The Sidrauski optimizing model can be used to answer the question posed as the optimum quantity of money: What steady-state rate of inflation maximizes steady-state welfare? The answer, provided by Friedman (1969), is: that rate of inflation which makes the nominal interest rate zero. This usually negative rate of inflation satiates money holders with real balances which, since money is superneutral, are costless to produce; it may be necessary to impose lump sum taxes to reduce the money stock at the required rate. For a variety of practical reasons, Friedman does not follow through on the logic of his argument by proposing that policy go so far as to aim to achieve the necessary rate of deflation.

A frequent objection to the optimum quantity of money conclusion is that increases in real balances would displace capital accumulation, as occurs in some of the descriptive money and growth models. This displacement is difficult to demonstrate in certainty models in which individuals optimize over an infinite horizon. Tobin (1968) suggests using a model with overlapping generations and uncertainty in an attempt to begin to analyze the question of the effects of alternative government policies on capital accumulation. His interesting lead has not, to my knowledge, been followed up. The issue is closely related to that of the burden of the debt, in which it can be shown that the debt is a burden in life-cycle models but not in a model of optimization over an infinite horizon.

---

[10] These points are elucidated in papers by David Levhari and Patinkin, and Rudiger Dornbusch and Jacob Frenkel.

The major lesson of the optimum quantity of money literature, which is closely related to the much older literature viewing inflation as a tax on real balances, is that economists have little basis for discussing desirable rates of correctly anticipated inflation: the only cost of such inflation appears to be liquidity costs, which are small. An urgent order of business remains clarification and quantification of the costs of inflation discussed by Phelps (1972).

I conclude this section by mentioning a method of analysis—temporary equilibrium analysis—rather than a particular problem. The method is in fact typically used by monetary economists; equilibrium is established on existing markets conditional on expectations about the future and on stocks inherited from the past. A similar framework has recently been used by general equilibrium theorists to study problems of monetary theory, such as the quantity theory of money and the classical dichotomy (Jean-Michel Grandmont and Yves Younes, 1972) and the optimum quantity of money (Grandmont and Younes, 1973). The basic existence result in these studies is that equilibrium will not exist if expectations of future prices react too strongly to changes in current prices.

The ambiguity of most of the results discussed in the money and growth literature is distressing to the theorist, though no doubt cheering for the econometrician: numerical values of parameters do matter. Some conceptual problems have been clarified in this literature—for instance, the relationship between stock and flow equilibria (Josef May and Duncan Foley and Sidrauski). Problems requiring further research—price dynamics, expectations, and the micro foundations of macro models—become painfully clear in the development of the literature. None of these, however, is inherently related to growth.

IV. Rational Expectations

The dependence of the dynamics of the price level on the process of expectations formation is obvious from the literatures on money-and-growth and the new microeconomic foundations of inflation (Phelps 1970). In the money and growth literature, each assumption about expectations of inflation implies a different path for inflation. Although adaptive expectations have been widely used, explanations for their use are few: adaptive expectations seemed reasonable and proved useful in explaining data, as in Philip Cagan's pioneering work on hyperinflation. However, even in Cagan's model, adaptive expectations are systematically wrong in predicting, say, the behavior of the inflation rate in response to a change in the growth rate of money.

The rational expectations hypothesis is that the expectations of economic agents are the same as the predictions of the models in which they are assumed to be actors. Thus the hypothesis removes part of the excessive freedom of assumptions in dynamic models. In deterministic models, rational expectations are equivalent to perfect foresight, so that the expected path of prices in a model is the path that, according to the model, subsequently ensues, given the expectations. Perfect foresight in monetary growth models implies instability of the inflationary process if no jumps in the price level are permitted; alternatively, if jumps in the price level can take place, perfect foresight implies there are many possible paths for prices, only one of which is stable. This saddle-point indeterminacy-instability result is similar to the well-known Hahn paradox of capital theory and raises the question of whether the unique stable path will be followed in a private economy. William Brock presents an optimizing money and growth model in which the economy follows the stable-perfect foresight path because it is utility maximizing. The types of policy actions which are consistent with the perfect foresight assumption are limited: in perfect foresight models, preannounced changes in the growth rate of money lead to changes in the inflation rate before the change in the monetary growth rate occurs.

In a stochastic model, the rational expectations assumption is that the subjective probability distributions used in decision making by private agents are the same as the probability distributions for the relevant variables implied by the model. More generally, if individuals do not have all the information used in the model, expectations are rational if they are optimal predictions based on the information available to the individual.

Rational expectations in the context of limited information by market participants have been used by Robert Lucas (1972 and 1973) to generate a short-run Phillips curve in models in which the long-run Phillips curve is vertical. In Lucas' models, participants in a given market are aware of current conditions only in that market but do know economywide conditions for previous periods. Demand in a single market may change either because of real demand shifts or because of economywide changes in nominal demand caused, say, by monetary growth. Since suppliers in a single market cannot distinguish the cause of a current disturbance in their market, they do adjust output to nominal disturbances even though they would not do so if they were informed of the cause of the disturbance. Thus, in the short run, suppliers can be fooled into increasing output by a process which also increases prices, generating a short-run Phillips curve. However, the tradeoff vanishes over time as knowledge of conditions elsewhere in the economy is obtained.

Models in which the Phillips curve is a result only of expectational errors have radical implications for policy when combined with the rational expectations assumption—essentially because the latter causes long-run results to hold in the short run. In a series of papers, Thomas Sargent and Neil Wallace show that, under rational expectations and with an expectational Phillips curve, the behavior of the economy is invariant with respect to the preannounced systematic part of a monetary policy rule since, essentially, only surprises produce a real response. It is possible for the monetary authority to affect output, at least temporarily, by concealing its actions; John Taylor uses a model in which individuals form expectations rationally, but the use of a random monetary rule enables the monetary authority to fool individuals and so affect output. It is, of course, another matter as to whether the monetary authority should want to affect output. More generally, Lucas (1973) uses rational expectations as the basis for a forceful restatement of the position that the use of existing econometric models for policy evaluation is dangerous since the parameters of such models can be expected to change with policy.

Although the rational expectations hypothesis provides a consistent explanation of some cyclical phenomena, it arouses a number of strong reactions. One reaction is that it is absurd to assume individuals can make the necessary calculations since most economists still cannot do so; however, individuals and firms do read newspapers and buy consulting services, and their expectations are influenced by model predictions. The rational expectations assumption *is* excessively strong since it ignores uncertainty over model selection, but it is a more persuasive starting point than the assumption that a rule of thumb is used for expectations formation which is independent of the stochastic properties of the time path of the variable about which expectations are formed. A second reaction is not so much to rational expectations as to expectational Phillips curves; it is argued that there are real reasons for short-run price inflexibility which produce a short-run Phillips curve and that the policy implications of rational expectations mentioned in the preceding paragraph are accordingly incorrect. While I am in sympathy with this view, I think it fair to say that a convincing model underlying the argument has yet to be produced. A third reaction is to the perfect foresight models and the restrictive role for policy in them; while perfect foresight is an unattractive assumption the cure is not necessarily to use nonrational expectations assumptions, but instead to use nondeterministic models.

V. Conclusions

Summary conclusions on each of the four branches of the literature surveyed here and some comments on gaps which need filling are presented in the preceding sections. There are other areas in which theoretical research is required and is taking place: existing models of money demand are chronically unable to explain the volume of money holding even though they do successfully predict the direction of interest rate changes on demand; advances in the theory of finance are being incorporated into macro models; the effects of indexing on the operation of financial markets and economic stability are being studied; monetary theory is being incorporated into trade theory. However, it is clear from sections II–IV of this paper that many of the major outstanding questions in monetary theory, and macroeconomics more generally, are related to the Phillips curve (or the famous missing equation) and concern the short-run dynamics and costs of inflation and unemployment. It is in this area of dynamics that most progress has been made in the past quarter century, and it is in this area that progress is most needed.

REFERENCES

R. Barro and H. Grossman, "A General Disequilibrium Model of Income and Employment," *Amer. Econ. Rev.*, Mar. 1971, *61*, 82–93.

J.-P. Benassy, "Neo-Keynesian Disequilibrium Theory in a Monetary Economy," *CEPREMAP*, Paris 1974.

W. Brock, "Money and Growth: The Case of Long-Run Perfect Foresight," *Int. Econ. Rev.*, Oct. 1974, *15*, 750–777.

K. Brunner and A. Meltzer, "The Uses of Money: Money in the Theory of an Exchange Economy," *Amer. Econ. Rev.*, Dec. 1971, *61*, 784–805.

P. Cagan, "The Monetary Dynamics of Hyperinflation," in M. Friedman, ed., *Studies in the Quantity Theory of Money*, Chicago 1956.

R. Clower, "The Keynesian Counter-Revolution: A Theoretical Appraisal," in F. H. Hahn and F. P. Brechling, eds., *The Theory of Interest Rates*, New York 1965.

R. Dornbusch and J. Frenkel, "Inflation and Growth: Alternative Approaches," *J. Mon., Cred., Bank.*, Feb. 1973, *5*, 141–56.

J. Flemming, "The Consumption Function when Capital Markets are Imperfect: The Permanent Income Hypothesis Reconsidered," *Oxford Econ. Pap.*, July 1973, *25*, 160–72.

D. Foley and M. Sidrauski, "Portfolio Choice, Investment and Growth," *Amer. Econ. Rev.*, Mar. 1970, *60*, 44–63.

M. Friedman, *The Optimum Quantity of Money*, Chicago 1969.

———, "A Theoretical Framework for Monetary Analysis," *J. Pol. Econ.*, Mar./Apr. 1970, *78*, 193–238.

J.-M. Grandmont and Y. Younes, "On the Role of Money and the Existence of a

Monetary Equilibrium," *Rev. Econ. Stud.*, July 1972, *39*, 355–372.

——— and ———, "On the Efficiency of a Monetary Equilibrium," *Rev. Econ. Stud.*, Apr. 1973, *40*, 149–166.

H. Grossman, "Money, Interest, and Prices in Market Disequilibrium," *J. Pol. Econ.*, Sept./Oct. 1971, *79*, 943–961.

F. Hahn, "On the Foundations of Monetary Theory," in M. Parkin and A. R. Nobay, eds., *Essays in Modern Economics*, London 1973a.

———, "On Transaction Costs, Inessential Sequence Economies, and Money," *Rev. Econ. Stud.*, Oct. 1973b, *40*, 449–462.

W. Heller, "The Holding of Money Balances in Equilibrium," *J. Econ. Theory*, Jan. 1974, *7*, 93–108.

J. Hicks, *Critical Essays in Monetary Theory*, Oxford 1967.

K. Iwai, "The Firm in Uncertain Markets and Its Price, Wage and Employment Adjustment," *Rev. Econ. Stud.*, Apr. 1974, *41*, 257–276.

H. Johnson, "Monetary Theory and Policy," *Amer. Econ. Rev.*, June 1962, *52*, 335–384.

———, "Money in a Neo-Classical One Sector Growth Model," in *Essays in Monetary Economics*, 2d ed., Cambridge 1969.

R. Jones, "The Origin and Development of Media of Exchange," unpublished manuscript, 1974.

J. M. Keynes, *General Theory of Employment, Interest, and Money*, New York 1936.

M. Kurz, "Equilibrium with Transaction Cost and Money in a Single Market Exchange Economy," *J. Econ. Theory*, Apr. 1974, *7*, 418–452.

A. Leijonhufvud, *On Keynesian Economics and the Economics of Keynes*, New York 1968.

D. Levhari and D. Patinkin, "The Role of Money in a Simple Growth Model," *Amer. Econ. Rev.*, Sept. 1968, *58*, 713–753.

R. Lucas, "Expectations and the Neutrality of Money," *J. Econ. Theory*, Apr. 1972, *4*, 103–124.

———, "Some International Evidence on Output-Inflation Tradeoffs," *Amer. Econ. Rev.*, June 1973, *63*, 326–334.

———, "Econometric Policy Evaluation: A Critique," unpublished manuscript, 1973.

J. May, "Period Analysis and Continuous Analysis in Patinkin's Macroeconomic Model," *J. Econ. Theory*, Mar. 1970, *2*, 1–9.

F. Modigliani, "Liquidity Preference and the Theory of Interest and Money," *Econometrica*, 1944, *12*, 45–88.

J. Niehans, "Money and Barter in General Equilibrium with Transactions Costs," *Amer. Econ. Rev.*, Dec. 1971, *61*, 773–783.

J. Ostroy, "The Informational Efficiency of Monetary Exchange," *Amer. Econ. Rev.*, Sept. 1973, *53*, 597–610.

D. Patinkin, *Money, Interest and Prices*, 2d. ed., New York 1965, ch. 13.

E. Phelps, ed., *Microeconomic Foundations of Employment and Inflation Theory*, New York 1970.

———, *Inflation Policy and Unemployment Theory: The Cost-Benefit Approach to Monetary Planning*, New York 1972.

R. Radner, "Competitive Equilibrium Under Uncertainty," *Econometrica*, Jan. 1968, *36*, 31–58.

T. Sargent and N. Wallace, "'Rational Expectations', The Optimal Monetary Instrument, and the Optimal Money Supply Rule," unpublished manuscript, 1973.

M. Sidrauski, "Rational Choice and Patterns of Growth in a Monetary Economy," *Amer. Econ. Rev.*, May 1967a, *57*, 534–544.

———, "Inflation and Economic Growth," *J. Pol. Econ.*, Dec. 1967b, *75*, 796–810.

J. Taylor, "Monetary Policy During a Transition to Rational Expectations," unpublished manuscript, 1974.

J. Tobin, "Money and Economic Growth," *Econometrica*, Oct. 1965, *33*, 671–684.

———, "Notes on Optimal Monetary Growth," *J. Pol. Econ.*, July/Aug. 1968, *76*, 833–859.

Empirical Monetary Macroeconomics: What Have We Learned in the Last 25 Years?

By W. C. BRAINARD AND R. N. COOPER*

Monetary economics conveys the impression of great disagreement within the economics profession, and indeed the professional debates have often been heated. But behind the debates over policy there is a great deal of consensus on the importance of monetary variables for the working of national economies and on the mechanisms through which they exert their influence. The title of this session reminds us that twenty-five years ago this was not so. When Howard Ellis wrote "The Rediscovery of Money," the postwar revival had just begun. There was general skepticism about the ability of monetary policy to influence the economy, and the Oxford surveys were widely cited as the empirical basis for disbelief in the effect of monetary policy on investment. Despite the emergence in the intervening years of wide agreement about the importance of monetary policy, the empirical basis for many of our beliefs, and a fortiori for distinguishing among our differences, has remained weak.

In casually accepting our present assignment, we failed to appreciate just how complex the question posed in the title is. This is true even when the subject is confined, as here, to the monetary economics of the business cycle, leaving aside both microeconomic and steady-state growth considerations. What does it mean to say we have "learned" something? And to whom does the "we" refer?

We interpret the first question to refer to how much we now know, compared with twenty five years ago, about managing the economy with instruments of monetary policy. Empirical knowledge is embodied in reliable quantitative economic relationships. This perspective does not, of course, encompass all of monetary economics, but it does encompass most of the professionally salient issues.

Who are "we"? There are often marked differences in the "knowledge" embodied in the most advanced work, in conventional professional wisdom, in ordinary textbooks, and in monetary officials. To what group should we apply our test of knowledge? Certain individuals are at the forefront of the profession, some with a pack of followers, others in splendid isolation; some are recognized at once, others with a short lag, and still others drop into oblivion to be rediscovered much later. For example, Irving Fisher explored what we now know as the Phillips curve in 1926 with an econometric sophistication which rivals all but the most recent work in the area. Diligent search can usually find a precursor to any idea. As Roberta Wohlstetter discovered in her study of the Pearl Harbor attack, the problem is not the inadequacy of information; rather, it is to separate the true signal from the noise. In an active intellectual discipline, this is always done with much greater clarity long after discoveries are first made. For this reason, while it might seem most appropriate to concentrate our attention

---

* Yale University.

on the most advanced work, it is neither easy nor fully appropriate to do so, since we may not even know where the true frontiers of the subject are at the present time.

In some respects, monetary economics as a whole was just being "rediscovered" as an important field in the late 1940's and a twenty-five year survey starts near a low point. The outpouring of empirical work in monetary economics during the past two decades—and especially during the past decade—has been vast. A simple bibliography of professional articles alone would exhaust the space allotted to us. We must be highly selective and, therefore perforce, somewhat impressionistic. We first present some quantitative information on professional articles on empirical monetary economics in the late 1940's and in the early 1970's, along with a brief characterization of these articles. We then discuss the chain of monetary influence as embodied in empirical macro models of the *U.S.* economy. Finally, based on these comparisons and on an impressionistic sampling of other literature, we offer some generalizations about improvements in our knowledge and about areas where our ignorance remains great.

We are concerned mainly with the endpoints of our period. Limitations of space prevent consideration of many important milestones within the period, such as Gurley-Shaw, the Radcliffe Report, the Commission on Money and Credit, Friedman-Schwartz.

I. Quantitative Assessment

One measure of the relative importance and the character of empirical work in monetary economics now as compared to the late 1940's can be obtained by surveying articles in the professional journals. We have counted the articles (including notes and replies) on monetary economics in four journals, representative of the best current research by the profession and published at both the beginning and end of our period: the *American Economic Review*, the *Journal of Political Economy*, the *Quarterly Journal of Economics*, and the *Review of Economics and Statistics*. The results confirm the general impression of a great increase in work on monetary economics: in the period 1948–50 there were twenty-six articles in the four journals, of which fifteen were theoretical and eleven were empirical. In the period 1970–72, there were eighty-seven articles on monetary economics, of which fifty-one were theoretical and thirty-six were empirical. This undoubtedly understates the growth since at least one major journal, devoted entirely to research in the monetary area, was created during this period, the *Journal of Money, Credit and Banking*. Thus, there was a sharp increase in both theoretical and empirical work on monetary economics. Moreover, there was a marked shift in the character of the empirical articles: in the earlier period three were historical-analytical and eight involved graphical analysis of two variables.[1] None contained regressions. In the latter period, by contrast, thirty-three involved regression analysis, two involved simulations, and one contained only graphical analysis.[2]

This great increase in quantitative empirical work, of course, reflects the increased availability of better data and of high speed computers permitting their analysis, but also greater appreciation of the intrinsic complexity of monetary relationships, calling for multivariate and simultaneous equation analysis.

---

[1] Examples of empirical articles from 1948–50 are E. Marcus, Kenneth Roose, James Tobin, Clark Warburton, and Charles Whittlesey.

[2] These findings may have reflected in part a change in editorial policy regarding regressions. But regression analysis was used in other areas of economics, especially in the *Review of Economics and Statistics*. And *Econometrica* also contained no regression analysis on monetary economics during 1948–50.

In certain respects, the sharp growth in quantification itself represents a substantial gain. It compels more exact thought, both as regards the relationships among variables and as regards the variables themselves. Theoretical formulations can speak of "the" interest rate and "the" money supply, for example, but the quantifier must choose precisely what interest rate and what definition of the money supply he is going to use. This is a useful discipline, if only to indicate that our theoretical concepts are not quite as operationally sharp as we would like them to be. Second, quantification permits some assessment of the quantitative importance of the variables in question; effects of negligible practical importance often command disproportionate theoretical attention. On the other hand, a preoccupation with quantification may result in neglect of important but not easily measured variables. And quantification is itself no assurance against misspecification.

One of the most dramatic changes during the period is the increase in the number of working econometric models purporting to describe the behavior of the *U.S.* economy. No less than eleven models are participating in the National Bureau of Economic Research-National Science Foundation (*NBER-NSF*) continuing seminar on the comparison of econometric models, and since 1970 at least four different models have been used for regular quarterly forecasts one to five quarters ahead. Models of this type have the advantage of embodying concrete quantitative information about empirical linkages in the economy.

In 1950, econometric models were in their infancy. Few thought of them as a practical tool for forecasting the economy or the impact of policy; rather the concern was with the extent to which a model could capture the salient features of the economy and its cyclical behavior, and with problems of econometric methodology. Nevertheless, comparison of the difference in the treatment of the monetary sector in these models and their modern counterparts gives dramatic testimony to the cycle in the profession's perception of the importance of monetary factors in the determination of income and prices. As reflected in the models, the recovery of money from the Great Depression was protracted. When Marc Nerlove did his survey of the six models estimated from postwar data in 1966, for example, only two had more than two endogenous financial variables, and three had no monetary sector at all. This contrasts sharply with the remarkable model by Jan Tinbergen (1939), which contains an elaborate monetary sector. By the end of the period most of the models included a substantial monetary sector. Perhaps the most sophisticated and most widely emulated is the financial sector of the MIT-PENN-SSRC (*MPS*) model. In it monetary policy has powerful effects on income and works through a rich and complicated set of linkages. While it has been the subject of criticism by monetarists and nonmonetarists alike, it has gone a long way towards resolving some of their differences.

At the same time that structural models were in the process of reintroducing money, there was an effort, stimulated by Milton Friedman and David Meiselman, to determine whether simple single-equation Keynesian or monetarist models provide a better explanation of money income. In retrospect, it appears clear that it is not possible to discriminate between competing macro models simply on the basis of in-sample fit. Given the highly serially and cross-correlated nature of economic time series, the range of choice of variables for use in any model, and the flexibility in the specification of lags, marvelous fits of historic data can be obtained by models with widely different implications for the behavior of the economy and its response to policy. Out-of-sample forecasts provide a better test, but it appears difficult to discriminate between the simple Keynesian and monetarist models on this basis. Although out-of-sample forecasts do not provide a clear verdict on which of these single-equation models predicts best, they do show persuasively that these models, taken as a class or individually, cannot compete successfully with the more complicated models, structural or otherwise, nor for that matter with the median of judgmental forecasts. Today, most economists would probably agree with William Poole and Elinda Kornblith when they say (p. 915), "neither the simple Keynesian nor the simple quantity-theory models provide an adequate understanding of business cycle fluctuations . . . there simply is not much empirical content in the single-equation approach as employed in the studies examined here."

Given the fact that economic data do not distinguish sharply between competing views, the plausibility of theoretical explanation will continue to bear heavy weight in distinguishing among alternative hypotheses as well as in guiding our exploration of the data. For that reason, students of monetary economics will want to know how money works. They will want to know how specific components of demand are connected to money markets and the accuracy of predictions about the behavior of the links in that causal chain.

II. Channels of Monetary Policy

Earlier monetary writings are very rich in their verbal treatment of the influence of monetary variables on economic activity and at a general level our knowledge has not advanced beyond them. No one has discovered a brand new channel of influence, and indeed some of the monetary effects identified earlier, such as the importance of business and consumer expectations, or "confidence," have not yet been satisfactorily incorporated into quantitative models.

We will organize our discussion of linkages from the perspective of the policymaker. From his view, the instruments of policy are directly controlled. Thus, the direction of economic causation runs from his instruments to the rest of the economy. A simplified rendition of the chain of influence is as follows:

$$\begin{tikzcd}
& & & & P \\
CB \arrow[r, "(1)"'] & H \arrow[r, "(2)"'] & MM \arrow[r, "(3)"'] & Y \arrow[ur, "(4a)"] \arrow[dr, "(4b)"'] & \\
& & & & Q
\end{tikzcd}$$

where CB = the central bank, H = high-powered money, MM = money markets, Y = money income, Q = the volume of output, and P = the overall price level. The important empirical questions concern the magnitude and the reliability of each linkage. From the econometrician's point of view, of course, policy however measured is unlikely to be exogenous. This fact creates problems of estimation and interpretation of both reduced form and structural equations. Similarly, the influence of exogenous variables and the feedback from price and output greatly complicate the estimation of this chain.

In this scheme, what is the instrument of control? Taken literally, the central bank controls its portfolio of assets, the discount rate, reserve requirements, and other regulations which restrict and influence the behavior of banks and other institutions. But effective control may extend much further into the private realm. What effective control is, of course, depends upon the objectives of policy and the horizon in question.

Early discussion tended to focus on the quantity of currency plus demand deposits as the important element in MM. The central bank controlled this quantity, which was tied directly to Y by a velocity or demand for money function. Recently a number of studies suggest that control of M_1 over periods as short as a month may be difficult. Given other, longer lags in the system, it does not seem to us that this possibility is itself a cause for alarm. It is also apparent that over a cycle in economic activity there is a difference between controlling H and controlling money, even if money is defined narrowly. Earlier generations of students were trained in increasingly elaborate mechanics of the money multiplier, which tied the volume of bank deposit liabilities directly to the supply of high-powered money. The ability of the banking system to affect this multiplier by borrowing or altering excess reserves has long been recognized.[3] But while this slippage may decrease the magnitude of response to changes in H, it is not likely to be so rapid or so dramatic as to prevent control of a particular M over the medium term.

Over longer periods, postwar experience with the development of the federal funds market, the emergence of certificates of deposit, and the development of ties between American banks and the Eurodollar market demonstrate how institutional adaptation to profit-making opportunities and regulation may change the relationship between H and money, no matter how defined. Such endogenous institutional changes suggest that it may be difficult to specify any simple rule for monetary policy applicable to the long run.

It would be possible to tighten the linkage between H and a particular M by implementation of proposals such as that for 100 percent money (Lloyd Mints). However, the growth during the period in bank liabilities other than demand deposits, in deposit liabilities of financial intermediaries, and in other near monies makes clear the difficulty of settling on any single definition of money for policy purposes. Improving control over one variable is likely to increase slippages elsewhere in the system.

While there is still much disagreement about the best way to describe empirically the third linkage, connecting money markets and national output, monetary economists have come a long way towards agreement about general outline of the process. Such diverse writers as Tobin, Allan Meltzer and Karl Brunner, Friedman, and Franco Modigliani stress the importance of recognizing that a broad spectrum of assets and commodities are affected by monetary disturbances, and of distinguishing between the rates of interest, nominal or real, on financial assets and the required rate of return on capital or durable goods. This latter point, which can be found in Tinbergen but was neglected in virtually all models until the *MPS*, can be formulated in a variety of superficially different ways. According to all these authors, discrepancies between the required rate on physical assets and their marginal productivities, or equivalently, between the current valuation of these goods and their replacement costs, stimulate or discourage their production.

One important implication of this more complex view of the *modus operandi* of monetary policy is that the second and third linkages fuse, and changes in policy can influence Y without necessarily going through any particular monetary aggregate.

Traditional concern with the variability of "velocity" involves the third linkage. At a theoretical level it was long assumed that the essential issue was the interest elasticity of the demand for money, an empirical issue on which the relative effectiveness of monetary and fiscal policy was assumed to hinge. In models which incorporate a wide variety of financial markets, the role played by the demand for any particular definition of money is displaced by the demand for high-powered money, H. The now widely accepted view that the demand for money depends on interest rates then becomes only one reason for some elasticity in the demand for H.

---

[3] It is interesting to note that Tinbergen's model includes an explanation of bank demand for net free reserves not unlike that studied by A. J. Meigs and C. Morrison in the late 1960's and embodied in the *MPS* model thirty years later.

It is not difficult to construct models of financial and commodity markets which honor the distinctions among real and financial assets and rates, but the simultaneous adjustment of financial markets and imperfections in markets for existing physical assets make it difficult to utilize them econometrically. The builders of the *MPS* model, for example, did not find it possible to identify separate supply and demand equations for most of the wide range of financial assets represented in the model. Instead, they rely heavily on a "term structure" equation to tie the markets together. Among other things, this simplifying assumption ignores the cumulative impact of net government borrowing on the relative supplies of bonds, capital, and other assets. As a result, the effect of government debt on the required rates of return on capital cannot be captured in the model. This assumption also restricts the ability of monetary disturbances to affect the demand for real assets without going through a long chain of financial markets. This weakness is partially remedied in the recent Bosworth-Duesenberry model of flow of funds and other models currently under construction, which include a complete set of supply and demand equations for financial assets; but it seems unlikely that such models will be reliably estimated from time series data alone.

During the period under discussion an enormous amount of quantitative empirical work has been done on various sectors of the economy. This work suggests that monetary variables have a substantial effect on expenditures. Interest rates appear to have an important influence on expenditures for plant and equipment and for purchases by state and local governments. In the early 1950's much weight was put on the "availability" of funds, and we know that this is especially important today in influencing housing and perhaps other expenditures as well.

Another linkage between financial markets and expenditures arises from interest-induced changes in asset prices and from changes in the real value of financial wealth in terms of goods and services. In the *MPS* model, for example, interest rates can have a very large effect on stock market valuations and, even with small marginal propensities to consume capital gains, are responsible for making consumption very responsive to monetary policy. The magnitudes of these effects are open to question. We do not yet understand how much is permanent and how much is regarded by wealth holders as representing temporary vicissitudes in monetary policy, with negligible influence on spending.

There has been much debate over the fourth linkage, with Keynesians emphasizing the impact of increased money income on real output and monetarists emphasizing the impact on prices. Both schools of thought have acknowledged that both effects are generally present in the short run, and the discussion has come to focus on the relative magnitudes of the short-run effects and on whether any effect on output endures in the long run. Of the linkages identified above, this one is the least well understood; forecasting errors on price changes are large, we have virtually no policy instruments to influence the division between Q and P, and we have virtually no dynamic adjustment theory to indicate what division to expect at any moment of time. What empirical evidence we have suggests that we can forecast Q better than Y. That is, in the short run prices are influenced by factors not well captured by the models. Furthermore, the effect of changes in demand on prices appears to be slow.

The above chain of influence assumes, as most economists do, that the central bank is an autonomous agent and that it can determine H, high-powered money. Both assumptions are subject to question, at least under some important circumstances.

In democratic societies, central bank behavior is really endogenous to a model of the political process, so we need to close the circle of causation with a fifth linkage, running from P and Q to CB. This is more the domain of political scientists than of economists, but, as noted above, it affects the descriptive value of macro models and introduces simultaneous equation bias into econometric estimates. Moreover, it suggests that economists should not pretend that central banks can pursue freely their own preferences or those of their advisers, and should direct greater professional attention to enlarging and improving the range of instruments, nonmonetary as well as monetary, for influencing target variables.

The extent to which central banks can influence high-powered money can also be questioned. Work on the monetary approach to the balance of payments has reversed the whole chain of influence, at least for a small country with a fixed exchange rate. Q is determined by real factors, and P is determined by the exchange rate and the world price level. These together determine Y, which determines the demand for money, which in turn influences public behavior to assure adequate supply, if not directly from the central bank, then through the balance of payments. Attempts by the central bank to supply more or less than the desired amount of money will simply result in deficits or surpluses in the balance of payments. David Hume has been rediscovered, and the money supply becomes completely endogenous.

This line of reasoning is less applicable to a large country, to a country whose currency is a reserve currency, and to a world of flexible exchange rates. But it does not become wholly inapplicable under any of these circumstances as they are likely to obtain in reality. Moreover, the use of a national currency as an international currency in private transactions raises serious questions about operational meaning of M: are *foreign*-held dollar balances related to money income in the United States? Or in the country of the holder? Or both? The substantial size of the Eurocurrency market gives rise to a host of measurement problems which in turn reflect some looseness in our concepts of money.

III. Predictive Accuracy

Better knowledge should lead to better forecasts of the economy. Forecasting out-of-sample offers one of the few genuine tests of whether or not our knowledge of reality has improved and of the influence of policy actions on the economy. In fact, there seems to have been a substantial improvement in our ability to forecast short-term economic developments. The median forecast in the early 1950's had a mean error of about 2 percent in forecasting nominal GNP four quarters ahead (Victor Zarnowitz). Formal macroeconomic models had a mean error of only about 1 percent in such forecasts in the early 1970's (S. McNess). It is also clear, however, that forecasting performance is not a powerful tool for testing alternative model specifications. Although the more complicated models appear to have overtaken judgmental forecasts and on average outperform single-equation or extrapolative forecasts, it is not clear how important elaborate specification of structure is in achieving this improvement. A substantial amount of judgment is typically used in the operation of the more complicated models. Even so, Ray Fair's relatively simple 20-equation model, containing very little structure and used mechanically, performs nearly as well as the more complicated models and appears to outperform them when they are also used mechanically.

These short-term out-of-sample forecasts on aggregate economic variables do not test the model's ability to predict the response of the economy to major changes in policy. In the absence of experiments in policy, our only guide to performance in this regard is the ability of the models to predict well the large number of endogenous variables which transmit the effects of policy actions to the target variables. Unfortunately the record so far is not outstanding (Charles R. Nelson).

One of the problems which plagues macro-model builders is the difficulty of distinguishing adjustment lags, expectations, and serial correlation of errors. These phenomena lead to a high degree of autocorrelation in many economic series, which is one of the reasons that naive autoregressive predictions do so well in short-term forecasting. They presumably are also the reason that it is typically quite difficult to pass tests of causation which call for strong relationships between movements in economic variables purified of autocorrelation. (See Christopher A. Sims 1972 and C. W. J. Granger.)

Treatment of lags in the early models was very primitive. Part of this reflected the absence of data on a more frequent than annual basis, and it seems clear that annual data do not provide very much information about the lags in most economic relationships. Quarterly models, starting with James Duesenberry, Otto Eckstein, and Gary Fromm, provide a substantial improvement in this respect, but confidence in the reliability of estimated lag structures is still quite low. In addition to all the other difficulties in estimating lag structures, recent work by Sims (1971) and J. Geweke suggests that, due to "stroboscopic" effects, estimation of continuous processes by the use of discrete time observations may yield a fitted lag structure very different from the true lag structure.

A closely related weakness in the models—and in our empirical knowledge generally—concerns the formation and the alteration of expectations. The frequent and convenient use of adaptive expectations implicitly assumes nonrational behavior on the part of the economic actors. Over time, rational expectations will take into account the actions of the policymakers, and this, in turn, can lead to an alteration of the structure of the economy itself, rendering any given model obsolete. Policy-induced institutional change, which has been very important in the *U.S.* financial sector over the past two decades, is one example of this; regulations create incentives to evade them. Another example is the alleged decline in money illusion and a tendency for investors and savers alike to respond to real rather than nominal interest rates. Under the circumstances of evasive changes in expectations and in economic structure, policymakers might well find that an optimizing strategy would involve deliberately introducing some ambiguity—for example, through randomizing—into their countercyclical actions. Close students of monetary policy might even say that the Federal Reserve has been using this principle from game theory all along.

REFERENCES

A. Ando and F. Modigliani, "Econometric Analysis of Stabilization Policies," *Amer. Econ. Rev. Proc.*, May 1969, *49*, 296–314.

B. Bosworth and J. S. Duesenberry, "A Flow of Funds Model and Its Implications," *Issues in Federal Debt Management*, Boston 1973, 39–149.

F. De Leeuw and E. Gramlich, "The Channels of Monetary Policy: A Further Report on the FRB-MIT Model," *J. of Fin.*, May 1969, 265–90.

J. Duesenberry, O. Eckstein, and G. Fromm, "A Simulation of the U.S. Economy in Recession," *Econometrica*, Oct. 1960.

R. Fair, *A Short-run Forecasting Model of the U.S. Economy*, Lexington 1971.

M. Friedman and D. Meiselman, "The Relative Stability of Monetary Velocity and the Investment Multiplier in the United States, 1897–1958," Commission on Money and Credit, *Stabilization Policies*.

C. W. J. Granger, "Investment Causal Relations by Econometric Models and Cross-spectral Methods," *Econometrica*, July 1969, *37*.

J. Geweke, "Temporal Aggregation in the Multivariate Regression Model," unpublished, 1974.

S. McNess, "The Predictive Accuracy of Econometric Forecasts," *New England Econ. Rev.*, Sept./Oct. 1973, 29–34.

E. Marcus, "The Interest-Rate Structure," *Rev. of Econ. and Statist.*, Aug. 1948, *30*, 223–26.

A. J. Meigs, *Free Reserves and the Money Supply*, Chicago 1962.

G. Morrison, *Liquidity Preferences of Commercial Banks*, Chicago 1966.

L. Mints, *Monetary Stability for a Competitive Society*, New York 1950.

C. R. Nelson, "The Prediction Performance of the FRB-MIT-PENN Model of the U.S. Economy," *Amer. Econ. Rev.*, Dec. 1972, *52*, 908–17.

M. Nerlove, "A Tabular Survey of Macro-Econometric Models," *Int. Econ. Rev.*, May 1969.

W. Poole and E. Kornblith, "The Friedman-Meiselman CMC Paper: New Evidence on an Old Controversy," *Amer. Econ. Rev.*, Dec. 1973, *53*, 908–17.

K. Roose, "Federal Reserve Policy and the Recession of 1937–38," *Rev. of Econ. and Statist.*, May 1950, *32*, 177–83.

C. Sims, "Discrete Approximations to Continuous Time Distributed Lags in Econometrics," *Econometrica*, 1971, *39*.

———, "Money, Income and Causality," *Amer. Econ. Rev.*, Sept. 1972, *52*, 540–52.

J. Tinbergen, *Business Cycles in the United States of America 1919–1932*, Geneva 1939.

J. Tobin, "Rejoinder," *Rev. of Econ. and Statist.*, Nov. 1948, *30*, 314–17.

C. Warburton, "Monetary Policy and Monetary Velocity," *Rev. of Econ. and Statist.*, Nov. 1948, *30*, 304–14.

———, "The Secular Trend in Monetary Velocity," *Quarterly J. of Econ.*, Feb. 1949, *63*, 68–91.

———, "The Theory of Turning Points in Business Fluctuations," *Quarterly J. of Econ.*, Nov. 1950, *64*, 525–49.

C. Whittlesey, "Memorandum on the Stability of Demand Deposits," *Amer. Econ. Rev.*, Dec. 1949, *39*, 1192–1203.

V. Zarnowitz, *An Appraisal of Short-term Economic Forecasts*, New York 1967.

DISCUSSION

MILTON FRIEDMAN, University of Chicago: The papers by Stanley Fischer and by William Brainard and Richard Cooper are excellent summaries of recent work in the respective areas each surveys and I have only minor quarrels with either paper. Yet neither seems to me to give a satisfactory answer to the question posed for this session. The trees almost completely obscure the forest.

Would it be apparent to anyone reading these papers that twenty-five years ago the bulk of professional economists regarded "money" as simply a five letter word describing an utterly passive magnitude that responded to other economic forces but had negligible independent influence; that they took it for granted that inflation and unemployment alike could be relegated to a benighted historical past, if only policyholders could be persuaded to adopt the new Keynesian insights about fiscal policy; that the only role assigned to monetary policy in this happy scenario was to keep interest rates low; and that the quantity theory was of interest only to historians?

The past twenty-five years have seen a revolution in professional thinking. The revolution is well described in the title of this session as a "rediscovery," for indeed, as I shall note shortly, it is far harder to specify what "we"—meaning the profession of economics—have learned in the past two hundred years than what we have learned in the past twenty-five years.

To try to delineate the forest, the major components of that rediscovery have been:

1) The interest rate is the price of credit, not of money. The price of money is the inverse of the price level. Monetary authorities can control directly the quantity of money and thereby determine the long-term rate of inflation.
2) Nominal interest rates must be sharply distinguished from real interest rates. One implication is that while more rapid monetary growth may lower interest rates in the short run, it will raise them in the long run as it produces inflation which is embedded in interest rates.
3) The price level and inflation are monetary phenomena, not institutionally determined data to be analyzed by psychologists and industrial relations and industrial organization specialists.
4) There is no stable Phillips curve tradeoff. This is the empirical counterpart of the critical role in monetary theory today of the distinction between "anticipated" and "unanticipated" magnitudes.

I wish I could say that this rediscovery reflects the influence of the scientific work in monetary theory and empirical analysis that was initiated by a small group of dissenters from the prevailing orthodoxy twenty-five years ago, a trickle that has since grown to a veritable flood, as Brainard and Cooper point out. But that is not the case. The rediscovery reflects primarily the impact of experience. How can any teacher of elementary economics today stand in front of a class and begin his exposition of what is inelegantly called "macroeconomics" with the statement: "throughout this course we shall treat the price level as fixed"? Yet that is what essentially every teacher of elementary economics told his students twenty-five years ago. Or how can he, with a straight face, draw a liquidity preference diagram with nominal quantity of money on one axis and nominal interest rate on the other and explain to his students in these terms why the prime interest rate today is in double-digits? Double-digit inflation and double-digit interest rates, not the elegance of theoretical reasoning or the overwhelming persuasiveness of serried masses of statistics massaged through modern computers, explain the rediscovery of money.

That it is a "rediscovery" is easily demonstrated. Re the interest rate and the quantity of money, David Hume wrote more than two hundred years ago, "Lowness of interest is generally ascribed to plenty of money [and so it was twenty-five years ago]. But money, however plentiful, has no other effect, *if fixed*, than to raise the price of labour. . . . In the progress toward these changes, the augmentation [in the quantity of money] may have some influence, by exciting industry, but after the prices are settled, . . . it has no manner of influence."

Re money, inflation and employment, that is, the Phillips curve issue, listen again to Hume: "Though the high price of commodities be a necessary consequence of the increase of gold and silver, yet it follows not immediately upon that increase; but some time is required before the money circulates through the whole state. . . . At first, no alteration is perceived; by degrees the price rises, first of one commodity, then of another; till the whole at last reaches a just proportion with the new quantity of specie. . . . In my opinion, it is only in this interval or intermediate situation, between the acquisition of money and rise of prices, that the increasing quantity of gold and silver is favourable to industry. . . . From the whole of this reasoning we may conclude, that it is of no manner of consequence, with regard to the domestic happiness of a state, whether money be in a greater or less quantity. The good policy of the magistrate consists only in keeping it, if possible, still increasing; because, by that means, he keeps alive a spirit of industry in the nation. . . . There is always an interval before matters be adjusted to their new situation; and this interval is as pernicious to industry, when gold and silver are diminishing, as it is advantageous when these metals are increasing."

You can see why I say that it is far harder to answer the question, "What have we learned in the past 200 years?" than "What have we relearned in the past twenty-five years?"

As I see it, we have advanced beyond Hume in two respects only: first, we now have a more secure grasp on the quantitative magnitudes involved; second, we have gone one derivative beyond Hume.

On the first point, while we clearly have advanced, we must not overrate the advance. I was taken aback this summer in rereading William Stanley Jevons to come across the sentence in an essay first published in 1863: "An expansion of the currency occurs one or two years previous to a rise in prices." I thought I had been the one who had identified the lead of money over output and nominal income as six to nine months and over prices as two years! And there was Jevons a century ahead of me.

On the second point, Hume took it for granted, as did Irving Fisher fifty years ago, that stable prices were the norm, that generally rising or falling prices would be a departure from the norm, so would be unanticipated and so be "advantageous" or "pernicious" to industry. Today we have been forced to recognize that a changing price level—empirically, a rising price level—may be embodied in the public's expectations so that it takes an acceleration or deceleration of monetary growth to have the effects that Hume attributed to a rising or falling price level. We may soon be forced to still higher derivatives.

The need to go one derivative beyond Hume is what underlies one of the most important recent strands of monetary theory, stressed alike by Fischer and Brainard and Cooper—namely, analysis of the formation of anticipations and particularly rational expectations.

Once we recognize, as Hume did, that only unanticipated changes in the quantity of money will systematically produce a divergence between the actual rate of unemployment and the "natural" rate, the way anticipations are formed becomes a vital issue in the evaluation of monetary policy. The recent work in this area discussed by Fischer, particularly by R. Lucas and T. Sargent, brings us back to Henry Simons. As Fischer noted, if anticipations are formed rationally, no fixed policy on the part of the monetary authorities will produce a divergence between the actual and "natural" rates of unemployment. Only continuously "fooling" the public will do that. Is it not far better, along with Simons, to get instead the cooperation of the public by announcing in advance a policy and sticking to it? Though Lucas and Sargent have stated the point formally far better than it has been stated earlier, this has always been, as I see it, an essential foundation of the preference by Simons and his followers, including myself, for a fixed and announced rule of monetary policy.

The need to go to a higher derivative also underlies, I believe, the contrast referred to by Brainard and Cooper between the excellent forecasting performance of various models in predicting nominal income and their extremely poor performance in predicting the breakdown between prices and output. The change in nominal income has been linked now as in the past to the first derivative of money through a relatively stable velocity function. But the division between price and output has been linked to the higher derivatives through their effect on the division of monetary changes between anticipated and unanticipated changes.

In this connection, I may perhaps close on a somewhat more optimistic note for theory—though not for our near-term future—by disagreeing with Brainard's and Cooper's statement that "we have virtually no dynamic adjustment theory to indicate what division to expect at any moment of time" between output and price changes. I believe we have a reasonably good dynamic theory—what we lack for policy purposes is not the theory but the political capacity to use the theory effectively.

That dynamic theory—which I shall spell out oversimply, using *U.S.* quantities to make it highly specific—is little more than an expanded version of Hume's, but one that uses the quantitative knowledge that we have accumulated. At any time, the effect of monetary change depends on the public's anticipations about monetary change. Start from a hypothetical, reasonably balanced situation when monetary growth has been proceeding for some time at a constant rate so that the public in general has adjusted to that rate. *GNP* in nominal terms will then be growing at about the same percentage rate as M_2, prices at about 3.0 to 4.0 percentage points less. Let the growth rate of M_2 accelerate. For something like six months, the main effect will be that actual balances will exceed desired balances, which may temporarily depress short-term interest rates but will have little other effect. After about six to nine months, the rate of growth of nominal *GNP* will accelerate, as holders of the excess cash seek to dispose of it. The increased spending, by precisely the process Hume described, will "excite industry," as producers facing unexpectedly high nominal demands treat the increase as special to them and so seek to expand output. For a time they can do so, because their suppliers too, including laborers, take the increase in demand as special and temporary and do not alter their anticipations. This, if you will, is the temporary Keynesian phase, where output responds more quickly than prices. In its course, prices do respond, rising more rapidly than before, and interest rates stop falling and start to rise. But it takes about eighteen months after output starts to quicken—or two years after money accelerates—for the main effect to have shifted from output to prices. During this period, anticipations are changing, reflected most sensitively perhaps in interest rates, but even after prices have started to absorb the bulk of the acceleration in money, anticipations have not fully caught up. In the next year or so they will, which will force a decline in the rate of growth of output back to or below the "natural level," producing the stagflation stage.

I hasten to add, first, that this is the briefest of sketches; second, that the numerical magnitudes are based on the *U.S.* experience today and will change. In particular, the lags will shorten if the fluctuations in monetary growth continue to become more erratic, as I fear they may. But to stick my neck out, rather than shelter it, I may say that I believe we are in the early stages of an episode of roughly the kind I have described. As I read the tea leaves that emanate from the Federal Reserve—as an aside, why do they make us read tea leaves, instead of following the ancient wisdom of Simons and the new wisdom of Lucas and Sargent and telling us clearly and simply what they plan to do?—monetary growth sharply accelerated in October. If this accelerated rate continues (roughly 10 to 12 percent for M_2), the current recession will end in some six months or so, but inflation will continue to taper off under the influence of an earlier deceleration of money to something like 6 or 7 percent, but then, by say early 1976, the monetary acceleration will bring the tapering off of inflation to an end; by mid or late 1976, the rate of inflation will be back to perhaps 3 percentage points under whatever turns out to be the rate of growth to which M_2 settles down from October 1974 on. Inflation will then accelerate further until about mid-1977 when, if monetary growth has not accelerated still further, we shall be entering into the next episode of stagflation. Needless to say, the details of this scenario may be distorted or concealed by price and wage control, but not fundamentally changed.

If this scenario does develop and we end up in 1977 with 10 to 15 percent inflation and growing unemployment, it will not be because we do not have a "dynamic adjustment" theory, but because we do not have the political will to apply the knowledge we have.

FRANCO MODIGLIANI, Massachusetts Institute of Technology: We must be thankful to Stanley Fischer for his excellent survey of major recent developments in monetary theory, and to William C. Brainard and Richard N. Cooper (hereafter B–C) for tackling, with a good measure of success, the difficult task of surveying developments in our understanding of the role of money and monetary policy in the determination of real and money output. Still, I am inclined to agree with Milton Friedman that the two papers have failed to convey fully the dramatic revaluation of the importance of money which has occurred in the economic profession since the Second World War. At the same time, I differ with him on two main issues: (1) I disagree with the view that the revaluation resulted primarily from the rediscovery of old truths, and (2) I have serious questions about his list of propositions rediscovered.

Dealing first with (2), I would hope that economists have *not* rediscovered half-truths, such as that the price of money is the reciprocal of the price level, but continue instead to remember that money has as many prices as there are commodities, current and future, for which it can be exchanged. Again, while agreeing that the recognition of real rates as a determinant of real and financial investment behavior is an important rediscovery, I suggest that it has occurred simply because inflation has made the distinction between money and real rates become empirically important. At the same time, I would stress the equally important discovery (and not rediscovery) that because of institutional forces—such as the extensive reliance on level-payment mortgages to finance acquisition of dwellings and the reliance on financial intermediaries to supply those mortgages, subject to ceilings on nominal rates—investment behavior can be affected also by nominal interest rates. This proposition is well illustrated by this year's collapse in the housing market.

Finally, I hope and wish that no one is rediscovering that inflation is money determined, because in my view it is again at best a highly oversimplified and misleading half-truth. It is agreed that a high rate of inflation could not long be sustained without a sufficiently fast rate of growth of M and that a sustained high rate of growth of M (relative to the growth trend of output) could not fail to be accompanied, eventually, by a high rate of inflation; but this is quite different from saying that every instance of inflation is caused by excessive money growth. The first appearance of inflationary developments is typically the result of an excess of demand relative to capacity, and this may be due either to overexpansive *monetary* or *fiscal policy*. Indeed, when I look at the recent inflationary experience going back to 1966, I find it hard to decide whether the prize for misbehavior should go to our fiscal or to our monetary policymakers! However, an inflationary development can also be due to exogenous cost-push, although in the case of the United States this factor has seldom played a significant role before 1973–74.

If the excess demand is maintained, then the inflation will continue, possibly at an accelerating rate, but it is correct to say that it is money determined only if it is maintained through an overexpansionary monetary policy. On the other hand, once the inflationary process gets started—whether through excessive demand or exogenous cost-push—it will take a long time to die down, even after the excess demand is eliminated. The reasons for this long lag in the response of wages and prices to demand conditions have become fairly well understood thanks to recent developments in the analysis of the mechanism of wage and price determination (search theory of the labor markets, oligopolistic markup theory), although for lack of space these developments were largely ignored or underplayed in the two survey papers. In this phase, the inflation is no longer due to excess demand-pull, but rather to what may be labeled endogenous cost-push and is clearly not money determined. This syndrome is again well exemplified by the current situation in which a severely overrestrictive monetary policy early in the year reduced demand way below capacity with unemployment around 7 percent and yet inflation continued at a two-digit rate. In this situation, in order to maintain demand and employment at a level generating enough slack to insure the gradual elimination of the inflation, *but no more*, it may be necessary for M to rise at a fairly fast clip. But it would be wrong to regard such a fast rise as the cause of the inflation, and more nearly correct to regard it as the consequence of the endogenous cost-push process.

This point might be elucidated by recourse to a parallel which I have used before. Consider the situation of a car being driven on a flat road with a posted speed limit by a hurried driver who is both law abiding and bent on conserving gas. His optimum policy is to go at the speed limit. There is some position of the gas pedal which will result in this speed. If, by error, the driver pushes the pedal beyond this "warranted" position, the car will exceed the speed limit—the analogue of demand-pull. The correct response is to reduce the pressure on the pedal at least back to the warranted position, but not to use the brake as this would be wasteful of gas. But even after the pedal is returned to the proper position, the car will continue for a while to exceed the speed limit—the analogue of endogenous cost-push—though it will eventually get back on course. And since the wheels are connected with the motor, while the car is slowing down the motor will be running too fast; but it is the wheels that push the motor, not the motor that pushes the wheels.

Turning now to the other area of disagreement, I share B–C's view that the revaluation of the role of money reflects not rediscovery, but rather the notable strides we have made in peering into the monetarists' "black box" and understanding the complex mechanisms that link the tools of monetary policy to aggregate real output, prices, and money output. Since everybody here has felt free to choose his own historical span of comparison—Fischer 10 years, Friedman 200—I may perhaps be allowed to go back 30 years to my 1944 *Econometrica* article on liquidity preference and the theory of interest and money. There are primarily two areas in which I note an enormous improvement in our understanding between that article and the present state of the art as evidenced, for instance, by the theory and empirical estimates embodied in the MIT-PENN-SSRC (*MPS*) model. The first is that of the wage-price mechanism, to which I have already referred. In the 1944 article, the wage was partly exogenous, partly in the nature of a ratchet, while the price level was determined by money income and a Marshallian competitive supply function, implying a counterfactual monotonic positive association between real wages and rate of unemployment. The *MPS* model of oligopolistic markup on minimum long-run average cost is theoretically consistent with and can account far better for the observed real wage behavior. It also implies—in contrast to some monetarists' views which even B–C seem to have implicitly embraced—that a change in money can in general affect prices and money income only by *first* affecting real output.

The second area of improvement is the transmission mechanism through which M affects real demand. In the original article, this mechanism consisted in hypothesizing that *the* money interest rates affected net investment, but with no clear notion of the underlying relation between stocks and flows.

The *MPS*, by contrast, includes a number of different mechanisms for different components of investments, and they are generally clearly spelled out. They include the James Tobin and D. Foley-M. Sidrauski putty-putty model summarized by B–C, as well as the Charles Bishoff putty-clay model, in which the cost of capital determines the capital intensity of gross additions to capacity. Furthermore, just about every component of aggregate demand (except exports and inventory investment) responds to interest rates. In particular, consumption responds through the effect on wealth of capital gains and losses due to changes in capitalization rates. But note that, despite this multiplicity of channels, the *MPS* would probably still be viewed as a nonmonetarists' model because, as in the 1944 model, money does not affect demand directly, but only through the vehicle of the family of interest and capitalization rates.

Having indicated the major areas of progress, I should mention two major aspects of the linkage mechanisms which are in great need of further development. One, stressed also by B–C, is the modeling of expectations. The other is the modeling of credit rationing (which may not be too important for the United States, but is central in many other systems).

Let me conclude by noting that the revaluation of the role of money has also resulted in a notable reduction in the cleavage of views concerning the importance and workings of money—largely, but not entirely, through the nonmonetarists moving toward the monetarists. Indeed, there is reason to believe that at present monetarists and nonmonetarists largely agree, at least qualitatively, about both the *short-* and *long-run* effects of monetary and fiscal policy. But this agreement may be somewhat misleading. For in the monetarists' view, the *long run* is around the corner and, therefore, what happens in the fleeting short run is of little relevance. By contrast, for the nonmonetarists it is the long run which is irrelevant because it is far away, and the short-run response is the only one that matters. It is this difference—plus a related difference in assessing the quantitative impact on demand of changes in exogenous variables other than monetary (and fiscal) ones—which, despite broad agreement of principle, would still lead, say, Friedman and myself to advocate much of the time, including right now, very different monetary and fiscal policies.

A CRITICAL LOOK AT THE KEYNESIAN MODEL: THEORY AND APPLICATION

The Coordination of Economic Activities: A Keynesian Perspective

By Robert Clower and Axel Leijonhufvud*

Almost four decades have elapsed since the publication of Keynes's *General Theory of Employment, Interest and Money*. In the interim we have learned, or at least we have been told, a great deal about "what Keynes said"—and also about "what Keynes meant," "what Keynes *really* meant," and, most recently, "what Keynes really meant, *really*." Still the debate continues. If past experience is any guide, indeed, numerous questions about Keynes's theory—including some of considerable intellectual and historical interest—will never be settled. Rather than add further tidbits to what looks suspiciously like an infinite regress, therefore, we propose to address ourselves on this occasion to the other aspect of the Keynesian debate that is mentioned in the title of this session: the model, or rather family of models, that economists have developed in response to the stimulus of Keynes's great work—"*IS-LM*, and all that."

What does "A Critical Look at the Keynesian Model" reveal? It seems to us that what it reveals most clearly is that *the Keynesian model produces useful and sensible conclusions only in such measure as the user of the model is a good economist.* After all, the basic textbook model imposes next to no constraints on what may be argued from it; worse yet, it is open to *ad hoc* modification and common sense extension in numerous directions at the whim of its user. As a consequence, virtually any bundle of policy measures may be advocated in "Keynesian" language and, in recent years, almost all policy discussions have been couched in just such terms. To state the case succinctly, the Keynesian model imposes virtually no analytical discipline upon its users and thereby grants them essentially unrestricted analytical license. So it is not surprising that the debate over Keynes's contribution, and over the extent to which the latter-day Keynesian model preserves it, should continue. But there is another and even more fundamental reason why the debate continues, namely, we have yet to resolve the central question posed by Keynes's assault on received doctrine: Is the *existing* economic system, in any *significant* sense, self-adjusting?

The standard Keynesian model does not address this central question. It merely leaves room for us to analyze concrete problems under all sorts of alternative presumptions about what the answer might be, e.g., that the private economy never shows any endogenous tendency to return to equilibrium, or that it always tends promptly to restore itself to equilibrium following any disturbance. Moreover, the standard model hardly allows us even to frame the central question in a manner that would direct research onto a promising track. The comparative

---

* University of California at Los Angeles.

statics of *IS-LM* constructions has a well-demonstrated capacity for converting questions about adjustment performance (or, if you will, stability) into questions about sundry elasticities in the minds of economists.

These observations do not exhaust what might be said about the Keynesian model's capacity for mischief. But the point is simply that, in our judgment, the standard model is incapable of development in the directions to which the central question requires that we turn our attention. In effect, if we wish to say anything worthwhile about the central question, we must start by averting our eyes from the Keynesian model. Where do we proceed from there?

As of the present time, it appears to us that the question posed by Keynes remains unanswered largely because of the continued failure of economists to provide a coherent account of the manner in which production, consumption, and trading activities of individual economic agents are coordinated in theoretical economic systems that bear a family resemblance to economies of actual record. If this surmise is correct, then what is called for now is either a radical reformulation of existing analytical paradigms or—as would seem clearly preferable from almost any point of view—a fresh interpretation of established ideas. We shall pursue the second of these alternatives, although this procedure is calculated to supply less grist for existing analytical mills than food for further thought.

To provide an alternative frame for macro theory capable of addressing the coordination question is a formidable task that is unlikely to be solved except by the concerted efforts of many workers. It is also necessarily a long-run task. (Perhaps that is why it never gets done—so many macroeconomists follow Keynes in being almost equally responsive to the pressure of immediate policy issues and to their own long-run mortality.) But it seems quite clear to us where we have to start. A theory capable of describing system behavior as a temporal process, in or out of equilibrium, requires a prior account of *how trade is organized* in the system. Equilibrium, steady-state theory has managed pretty well without such an account. Macroeconomic theory cannot do so. Microeconomic theories of how business and household units behave—of how production and consumption decisions are made—when the system is *not* in equilibrium will have to be predicated on some such account. How we conceptualize the organization of trade, therefore, is a question with priority over numerous other theoretical questions some of which might otherwise be of more interest and of more promise as objects of theoretical ingenuity.

I.

Our point of departure is established, general, competitive analysis as set out, for example, in Kenneth J. Arrow and F. H. Hahn's recent account of the subject. It is customary to view this body of theory as an almost literal description of an idealized economy in which the notional economic plans of individual economic agents are costlessly coordinated by a central intelligence unit—the so-called auctioneer. This neo-Walrasian conception of economic organization is sharply at variance with that of its patron saint, Leon Walras. Walras regarded his work as an attempt *partially* to characterize equilibrium states of an ongoing economic system. To suppose that Walras conceived his analysis to constitute a *complete* description even of relevant equilibrium conditions would be an egregious error—as is indicated, among other things, by his explicit denial of any such notion in Lesson 35 of the *Elements*. In effect, modern neo-Walrasian theory starts with the solution to Walras' problem, but then proceeds to ask not whether that solution is acceptable in relation to the problem considered by Walras, but rather whether we can conceive of an economy that is *completely* characterized by equilibrium relations of the kind identified by Walras! The answer to this query—the problem which, so to speak, fits Walras' solution—is, of course, affirmative; for the auctioneer model of standard theory describes just such an economy.

That this particular type of model is unsuitable for analyzing any but *virtual* disequilibria is now generally acknowledged. The question then arises: How might we best proceed to modify standard theory so that it can be used to deal explicitly, at least in principle, with *real* disequilibria? On first thought, one's natural impulse is to eliminate the auctioneer and work instead with models in which individual economic agents engage in trade on a strictly "do it yourself" basis. No doubt that is the direction in which a definitive solution to the problem of economic coordination ultimately must be sought. As of the present time, however, and probably for the near future, second thoughts counsel a more conservative procedure. Specifically, it seems preferable temporarily to continue working with models that postulate the existence of a central coordinator of trading activity, but to jettison two other assumptions of standard theory: (1) the assumption that trades take place only at prices that ensure collective consistency of individual trading plans; and (2) the assumption that trades can be negotiated and executed at no cost to individuals or society in terms either of foregone leisure or scarce resources.

Let us suppose, then, that individual agents can engage in trade only by incurring implicit or explicit search and bargaining costs that are largely independent of quantities traded. On the basis of otherwise standard assumptions, it can then be shown that individuals will abstain from consumption and amass positive inventories of tradeable goods in order to avoid prohibitively heavy costs associated with small-lot trading at frequent intervals. It can also be shown that, in the absence of arrangements that permit individuals to trade in any desired amounts at dates of their own choosing, contacts between individuals will be so costly that large potential gains from trade will remain unexploited. In these circumstances, natural forces of greed and competition might plausibly be invoked to provide a rationale for the gradual emergence of merchant traders and organized markets. In keeping with our earlier decision to stick as closely as possible to received doctrine, however, we shall proceed instead by simply postulating the establishment by social contract of a central supermarket in which individual agents may execute at will pairwise trades of any good for any other good at going rates of exchange determined by a central *trade coordinator*—the counterpart in our discussion of the familiar *deus ex machina* of the neo-Walrasian auctioneer.

So that individuals may be assumed to be able to trade at dates and in amounts of their own choosing, we suppose that the trade coordinator starts with sizable inventories of all tradeable goods. So that operating expenses may be met, and aggregate inventories adjusted over time to conform with prevailing needs of trade, we suppose further that the trade coordinator charges a fee on each transaction or, what comes to the same thing, maintains different rates of exchange on trades of x for y than on trades of y for x. Finally, we suppose that rates of exchange established by the trade coordinator are governed by two basic requirements: (1) relative rates of exchange are varied with a view to maintaining average quantities traded at levels that will ensure positive holdings of all traded commodities at virtually every point in time; and (2) differentials between buying and selling rates are varied so as to aim at a zero long-run average level of net profit. The first requirement effectively asserts that the trade negotiator is ultimately responsive to prevailing forces of excess demand; the second is a rough device to ensure that the trade negotiator acts as if he (or she) had to compete for customers with other shopkeepers.

So much for the introduction to our story, brief though it may be. All we require at this point is a coherent mental portrait—in terms of ideas that are neither entirely unfamiliar nor arrantly offensive to common sense—of an economy in which trade at other than stationary equilibrium prices may be presumed to occur more or less routinely. The model just outlined seems to meet these specifications. Accordingly, we turn now to a quick—in fact, desperately quick—discussion of the existence and stability of equilibrium, which is followed by an even faster romp through related problems associated with the identification of possible and probable sources of coordination failure.

II.

The existence of equilibria in our supermarket model can be shown to follow from essentially the same assumptions as ensure the existence of equilibria in standard theory. Broadly speaking, equilibrium requires that rates of exchange be such that average rates of purchase and consumption by individual economic agents equal average rates of production and sale for each commodity, that average inventory holdings of individual agents and of the trade coordinator be maintained at constant levels over time, and that actual inventory holdings of the coordinator be nonnegative at every point in time. The familiar zero-excess demand conditions of standard theory may be viewed as a subset of these conditions, valid in the special case where sale and purchase flows are continuous and buying and selling rate differentials are arbitrarily small. In any equilibrium situation, "notional" and "active" demands and supplies of individual agents will, of course, coincide. Subject to certain obvious and minor changes in terminology, moreover, all equilibria will be Pareto optimal.

As far as equilibrium situations are concerned, then, our supermarket model may be viewed as a straightforward generalization of standard theory. As in standard theory, so in the present model: goods are traded directly for other goods; i.e., both models portray organized barter economies. In principle, the trade coordinator might deal in all kinds of commodities—employment contracts, futures contracts, leases, rentals, loans, insurance—as well as in spot trading of physical objects. In all cases we should expect the trade coordinator to impose standard *quid pro quo* requirements—i.e., Say's Principle would be strictly observed in all trades. In the case of contracts that call for future performance of specific acts (payment of interest, rent, provision of labor services, etc.) we should expect the coordinator to impose further *accountability conditions* to guard against conscious fraud or unintentional overcommitment by individual economic agents. In the nature of the case, conditions of this sort would be costly to enforce, and all such costs would have to be borne ultimately by individual economic agents. Accordingly, we should expect trade in any but short-term and easily monitored and enforced contracts to be severely limited. In this respect, the present model stands in sharp contrast to certain versions of standard theory.

III.

Much more could be said about equilibrium properties of our model, but here the preceding rough and ready observations must suffice. Dealing next with stability, we should begin by remarking that in the context of the present model this issue cannot be discussed in terms simply of trader reactions to price changes and relative speeds of adjustment of prices. The essential question is whether the trade coordinator, acting intelligently and on the basis of information derived from past experience, can manage most of the time to maintain effective control of all inventories so that no individual economic agent is ever denied for very long the right to trade commodities that the trade coordinator normally would handle as a matter of course.

Two considerations merit special notice as factors that would tend to ensure global stability. First—and in contrast with standard theory—price variations may be presumed to be governed not by predetermined velocity coefficients, but rather by reasoned and intelligent decisions of the trade coordinator. Except in circumstances where trader reactions to price variations are both erratic and violent, therefore, it should be possible for the trade coordinator to devise some strategy of price adjustment that would ensure stability. Second—and again in contrast with standard theory—quantities actually traded by individual agents can be directly restricted by the trade coordinator in cases where price adjustments alone do not ensure effective inventory control. Thus, even in exceptional circumstances where global stability might otherwise be a serious problem, departures from equilibrium should be limited in magnitude and bounded in duration. This is not to say that what Leijonhufvud has called "corridor phenomena" are *never* of any consequence; it is merely to say that such phenomena appear to be of no consequence in a world of organized barter.

IV.

The possibility of coordination failures obviously depends directly on the stability of disequilibrium adjustment processes and upon the nature and extent of external shocks to which the system is exposed. In our supermarket model, as in standard theory, it is possible to concoct cases in which everything goes wrong; the ideas of economists extend to many things that are not dreamed of in the real world. Unless one is chronically subject to acute bouts of intellectual perversity, however, such cases are not to be taken seriously. On almost any set of reasonable assumptions about the behavior of individual agents and the trade coordinator, coordination failures should be rare or nonexistent in a world of the kind portrayed by the model considered here. To be sure, if a substantial proportion of traded commodities were highly durable capital goods, then serious problems might arise from time to time in coordinating present production of such goods with future demand for their services (i.e., future purchases of consumption goods). Such problems might be avoided if appropriate facilities existed whereby present purchasers of capital goods could contract currently for future sale of consumer goods outputs; but such facilities would be extremely costly to operate, so it would seem highly implausible to suppose that they would in fact be provided by the trade coordinator. Though we cannot rule out stock-flow complications as a possible source of coordination failures, our own impressionistic feeling—Keynes to the contrary notwithstanding—is that this complication *alone* is not of overwhelming significance. More generally, we should argue that, in a world of organized barter in which the coordination of trading activities is completely centralized, there is no serious reason to object to the classical view of economic activity as a thoroughly harmonious concert conducted by Adam Smith's invisible hand.

V.

But suppose that some rudimentary monetary and credit complications are introduced into our model. Specifically, suppose that the trade coordinator decides to decentralize exchange activities by establishing distinct trading posts for all but one commodity and designates the exceptional commodity that is tradeable at all posts as *money*. We might then imagine coordination failures to occur from time to time because individual agents develop an irrational passion for cash rather than goods and services. But such failures—commercial crises, as they might be called by a student of eighteenth and nineteenth century economic history—would surely be short-lived, for real balance effects would operate with a vengeance in this kind of commodity-money world.

Suppose next, however, that trading specialists assigned to manage various trading posts were permitted to conclude trades on the basis of book credit as well as cash. Then we should expect a significant —and perhaps predominant—proportion of all trades to be made initially on the basis of book credit, cash settlement coming later in accordance with payment rules of the kind with which we are familiar in real life. In this situation, we should have to distinguish clearly between *means of payment*, which would consist of book credit *and* cash, and *means of settlement*, which would be represented by cash alone. In situations of disequilibrium, credit means of payment could expand or contract rapidly at the discretion of individual economic agents, virtually without reference to the short-run availability of means of settlement. Inflation of money prices could occur and be sustained for some time even with a constant stock of cash, and any major interruption in cash flows (an increasingly probable occurrence as means of settlement became an ever smaller fraction of total means of payment) could produce sharp restrictions on trade. This could lead, in turn, to rapid declines in output and employment, and to consequent price deflation, resulting eventually in widespread insolvency among trade specialists as well as bankruptcy of many individual economic agents. In this kind of world, corridor phenomena might well be of major consequence; i.e., sustained and serious coordination failures might occur because insolvency of trade specialists would temporarily eliminate from the economy market homeostats that are essential for effective coordination of the notional economic plans of individual agents.

Of course the world in which we live is not like this at all. But suppose that it were. Or, more plausibly, suppose that our monetary and credit version of the supermarket model were a reasonable first approximation to the world in which we live. The implications are pretty obvious—and not at all reassuring.

VI.

Here our story must end for now, although in truth it is hardly begun—for, of course, we have yet to regain a "Keynesian perspective." The theoretical conception of the economic system with which we break off (1) lacks a central information-processing and bill-collecting agency; (2) has, instead, middlemen trying to coordinate production and consumption activities in each output market separately; (3) makes the management of stocks of inventories essential to the coordination of these activities; and (4) has the system potentially subject to commercial crises associated with expansions and contractions of the volume of bank and nonbank commercial credit. All of this might be J. S. Mill or Alfred Marshall. At most, we might stretch our claim to being "with it" to include Knut Wicksell and Ralph Hawtrey.

So what becomes of our intention to provide food for further thought? If there is such a thing as stale food for fresh thought, we can only hope that some readers will derive nourishment from noting that (a) we started from a conceptualization of economic process lodged in "modern" general equilibrium theory, (b) proceeded to complicate matters in what we dare claim to be an obviously desirable and "realistic" direction, and (c) by that route ended up with Mill!

The Keynesian Revolution Reconsidered

By ROBERT EISNER*

It all seemed so simple once. A new generation of economists, nurtured "in the new ideas," had less difficulty "in escaping from the old ones." We understood the determinants of underemployment and knew what to do about it. But now the previously rearguard actions of separate groups of neoclassicists and monetarists have been seen by some as a full-fledged challenge to basic Keynesian theory. And at the policy level, professional as well as public bewilderment grows through months of increasing inflation *and* unemployment.

Whence emanated the "underemployment equilibrium" fondly exposited to a captive generation of students and all others who would listen? We all did see the truth of Keynes's denial of the second of the two "classical . . . postulates," that "the utility of the wage . . . is equal to the marginal disutility of . . . employment." But why? Was it rigid wages because of irrational or institutional constraints? Was it rigid wages due to perfectly rational choice in view of imperfect information, search costs, and inelasticity of expectations with regard to future wage offers? Or after all, is wage rigidity, whatever the reason, an explanation of "equilibrium" in the general level of prices and unnecessary to an explanation of underemployment?

It is this last that we can properly attribute to Keynes, not merely the "Keynesians," to evoke Axel Leijonhufvud's distinction—and I mean by Keynes, *The General Theory*, not necessarily the *Treatise* to which Leijonhufvud would also have us refer. To Keynes, wages were not perfectly flexible downward; we need not be detained as to the reason, whether Keynes's own explanation or the current, sophisticated views in terms of information and search costs. And it is also true that Keynes viewed the monetary authority as deficient in its efforts and methods to keep the quantity of money sufficient to minimize long-term rates of interest. But it is important to see, both in terms of theory and policy, that wage frictions and monetary restrictions, while both sufficient conditions, were not necessary for Keynesian unemployment. This relates rather to an aggregate of individual preferences which do not bring about current consumption of total product and the lack of an *intertemporal* price structure which can insure the production now of capital goods for the future in an amount equal to the gap between consumption and the total production that would be effected by full employment.

It has been customary to compartmentalize the deficiencies in effective demand. On the one hand, consumption, for a variety of reasons relating to individual preferences, would be both bounded and fairly inelastic over plausible ranges of rates of intertemporal substitution. Investment, or the demand for current production of capital goods, is constrained by two blades of a scissors: production possibilities of functions which underlie expected rates of return and the long-run rates of interest which may be viewed as a cost or rental price of capital or rate of capitalization of expected future returns. Looked at either way, the possible rate of investment in a free economy is also bounded, and this I would insist is Keynes and not merely "Keynesian," as Leijonhufvud argues.

---

* William R. Kenan Professor of Economics, Northwestern University, and research associate of the National Bureau of Economic Research. I am indebted for helpful comments to R. A. Gordon, Harry G. Johnson, and Robert M. Solow.

Implications of rational choice and some evidence (Eisner, 1963, 1971) substantiate the possibility of a positive liquidity trap. Past history, including a history of rate variability, risk, risk aversion and information and transaction costs, indicates that at any point of time the public may be ready at some positive rate of interest to turn all of its bonds or other income-earning assets into cash and indeed, contrary to Don Patinkin (p. 349), issue more promises to future payments—bonds—than are currently in existence in return for additional cash. The liquidity blade of the scissors, however, need not be set at a positive rate of return. It suffices for the theory of underemployment that the blade be nonnegative; and we can hardly suggest a willingness to lend or invest at negative rates of return as long as there is any nonwasting, noncostly asset such as money.

It is the other blade of the scissors that has been questioned, originally explicitly by Frank Knight, with antecedents going back at least to Eugen v. Böhm-Bawerk, later by "Keynesians" such as Robert Solow and James Tobin, by Dale Jorgenson implicitly in his self-named "neoclassical investment function," and most recently by Leijonhufvud. One can appeal to considerable if disputable econometric evidence that investment demand is relatively inelastic, over plausible ranges, to available rates of intertemporal substitution, where these should depend upon both nominal interest, or expected earnings rates, and expected rates of price change. One can resort to casual empiricism or *gedank*-experiments which might suggest that, at any available technology, *profitable* investment opportunities are finite, or recall that a capital addition must, to be profitable, not only add to gross product but must add more than its original cost. Or one may merely look to the total structure of Keynesian theory and recognize that, were there no problem of matching *attainable* rates of interest to *acceptable* rates of return, Keynesian underemployment would vanish in reasonably functioning markets. Saving in excess of the investment demand at full employment would simply enough bring about a fall in the rate of interest sufficient to raise investment demand, if not reduce saving, to the appropriate equilibrium.

Clearly, rigid or downwardly inflexible wages are a sufficient condition for persistent unemployment in the face of a decline in aggregate demand. But while wages may in fact not be perfectly flexible, Keynes's argument was that, even if they were, full employment would not be assured. Given unemployment, with a nominal quantity of money held constant or which does not decline in full proportion with wages and prices, real cash balances would rise, thus lowering the rate of interest and raising the real value of outstanding bonds. The lower rate of interest would increase investment demand, but because of the shape of the liquidity preference function and ultimately, if it ever came to that, the liquidity trap, and the limited interest elasticity of investment demand, this "Keynes effect" would prove insufficient. The higher real value of bonds could raise consumption by those who are net holders of bonds. But unless there were an outside debtor such as a government, for whose real obligations the public (out of money illusion) would have no concern, or which would have a different horizon or set of time preferences, the increased real value of bonds for net creditors would be offset by a corresponding increase in the real obligations of net debtors. As for cash itself, this would have to be or correspond to a net asset of the public, perhaps as "outside" or government money, and the government would have to follow a "monetary" policy which in effect, by holding the money supply constant or not letting it fall in proportion to prices, transfers real purchasing power to holders of cash.

The policy implications involved in all of this seemed clear for a long time to most economists. For one thing, keep interest rates as low as possible and have the central bank operate directly on the rates relevant to investment by buying *long*-term bonds. This would be a step in the right direction, but given the limits to how low the rate of interest could be pushed and given the limits to investment demand, further action in the way of fiscal policy to raise consumption might prove necessary, along with direct government participation to "socialize" investment.

An important implication of the theory and policy, sometimes implicit and sometimes explicit, was that "stabilization" meant maintenance of full employment. The appropriate policy was not to lop off booms but to fill in recessions. While Keynes was explicit on this last, a full consideration of the system, both as originally formulated and as we should see it more clearly now, reveals basic weaknesses and contradictions in conventional countercyclical policy.

With regard to monetary policy and the interest rate, countercyclical policy is inconsistent with full employment policy. For full employment, we want no monetary bar to maximum investment. This means that we want no shortage of liquidity, *current or expected*, to drive long-term interest rates above a minimum which might, with appropriate consumption and investment functions, be the natural rate. But every time the monetary authority curbs the money supply to fight an inflation or cut off a boom it raises interest rates. In so doing it makes all the less credible any policy of low long-term rates and vitiates its efforts to lower rates in the next recession.

While the self-cancelling effect of monetary shifts should have been implicit in original Keynesian formulations, difficulties in affecting consumption by countercyclical tax measures have become apparent with later formulations of the consumption function, particularly by Franco Modigliani and Richard Brumberg and by Milton Friedman (1957), and the ineffectiveness of the 1968 surcharges in combating inflation. If current consumption depends not merely on current income but on wealth and expected future income as well, a temporary increase in tax rates may have only a minor effect in curbing aggregate demand. And with regard to the investment function, we have come to see that the paths of the acceleration principle and, perhaps a fortiori, substitution effects involve fairly long, well-distributed lags and expectations of the profitability of investment which, however mercurial on occasion, are not always easily swayed by transitory considerations.

The common thread to difficulties that even a fully competitive system may have in returning to full employment, or in ever attaining it, stems from the often uncertain and often inconsistent ties between the present and the future. Thus interest rates may not fall far enough now because of expectations by many that they will be higher in the future. Workers will not accept lower wages now because they expect that if they wait or look longer they may find a better paying job in the future. Abandoning the fiction of tâtonnement and recontracting, sellers of goods and services may hold out in the expectation that if they wait their reservation prices will be met. (Of course, in a world of imperfect competition, a fall in demand may raise or lower prices or leave them the same, depending upon changes in elasticities of demand and the nature of marginal cost curves.) And there is no guarantee that expectations of future wants and future prices will determine consumption plans for the present and the future which will mesh with current and future investment.

Through all the recapitulations and revisions of theory, we should keep prominent the recognition that money is not omnipotent. Its supply and demand equations are undoubtedly more important than those for peanuts, but not in principle different. If the supply of peanuts is reduced, even to zero, there would be repercussions in other equations as the demand for substitutes is felt. Similarly, the supply of money, by any particular definition, can be reduced with a demand for (and creation of) money substitutes then increased. Since, as Robert Clower and others have reminded us, money may usefully be viewed as a factor of production, reducing its supply may well reduce real output precisely in association with rising costs, a matter to which we shall return.

And then tied to our recognition of long-run consumption functions, related to far more than current income, and investment functions which are inelastic, unstable, yet hard to affect, is the critical fact of lack of perfect information and futures markets which would meet that lack. The only pervasive futures markets that exist are in money but even these are almost always in nominal terms. Except on cost-plus government contracts (the significance of which, as John Kenneth Galbraith has eloquently reminded us, should not be overlooked) producers can rarely guarantee their profits by sales of futures to match spot purchases of capital, labor, and materials. The lack of such futures markets at every stage may lead sellers, producers, and buyers to hold back, in risk aversion and uncertainty, leaving demand and output deficient in comparison with what they might be if all markets, for present and future goods and services, could be cleared currently.

This suggests, by way of policy, fiscal measures to change intertemporal price expectations. Instead of varying taxes on income, which have a wealth effect almost certain to be in the same direction for both current and future consumption, government should vary tax rates on goods and services. In some periods tax rates may become negative, that is, turn into subsidies, in order to encourage expenditures. In other periods, tax rates would be raised in order to discourage expenditures. Changes would be pointedly temporary to encourage intertemporal substitution. Thus when demand is deficient, tax rates on goods and services can be reduced temporarily with an explicit warning that they are likely to be raised in the future when demand may prove too high. The difference between a 10 percent subsidy and a 10 percent tax, for example, may well entail a sufficient shift in intertemporal price expectations to have a substantial effect on current demand.

And instead of permanent equipment tax credits, investment allowances, or accelerated tax depreciation, government might much better institute high but variable tax and subsidy combinations on investment expenditures generally, including nonprofit institutions and state and local governmental units as well. Thus, in periods when investment expenditures are to be encouraged, subsidies might run not to 7 percent but to 14 or 28 percent of any expenditures above, say, depreciation charges or an average of the last three years' investment expenditures. Where investment expenditures are to be discouraged, the subsidy can be converted to a tax. Again, knowledge that taxes and subsidies are temporary would generate a low, possibly negative, elasticity of effective price expectations and major intertemporal substitution of investment expenditures.

Tax policy may also be used to affect directly the demand for labor services. Where employment is slack, payroll taxes can be cut or suspended, or even converted into subsidies to encourage employers to hire workers. Where inflationary bidding up of wages has taken place, payroll taxes can be raised, perhaps particularly on *increases* in payrolls beyond what they were previously. And all of this focuses directly on the root cause of Keynesian departures from (hopefully not inflationary) full employment equilibria: a lack of an appropriate set of intertemporal prices or price expectations.

By further way of policy, certain old prescriptions for full employment still offer a sound foundation. As Evsey Domar long ago suggested, a governmental "guarantee" of a steady growth of income, in part presumably through a steady growth in "autonomous" government expenditures, tends to generate the investment and consumption demand which brings the economy to full employment. Along with this should go a fairly steady growth of the money supply, consistent with both Friedman (1968) and Keynes.

But beyond that, recognition of public concern with inflation, if nothing else, dictates the formulation of policies to improve the inflation-unemployment tradeoff. There are many obvious things that could be said about firm government action to promote competition. Too often ignored are the mass and maze of government institutions, actions, and policies which contribute to inflation. These include tariffs, import quotas, price supports, resale price maintenance, petroleum prorationing, regulatory authorities that keep rates up, purchasing and hiring policies that raise prices not only for what government buys, but for all buyers in direct or indirect competition with government.

To the objection that elimination of barriers to competition may have a once-over effect of lowering prices but no long-run influence on the rate of inflation, one should raise the "ratchet effect." Inflation may stem as much from nonfalling as from rising prices. In any functioning economy, demands and supplies are perennially changing. Some prices must hence be expected to rise while others fall. If tariffs, regulators, or monopolists "protect" us from price declines, the average movement of all prices can only be upward.

On the positive side, government can move to improve markets in general, and notoriously imperfect labor markets in particular, by utilization of modern information-processing techniques to match jobs to workers, not only locally but anywhere in the nation. Planning, encouragement, training and retraining, and subsidization should be used to bring to full employment both the transiently unemployed and the stubborn hard core. Indeed, beyond usually defined full employment, much can be done to expand the voluntary labor force by facilitating the freeing of slaves of domestic labor and seeing to it that there is optimal investment in the human capital of all of our population, including women, minorities, and the historically disadvantaged.

While we have suggested sobering considerations regarding the efficacy of both monetary and fiscal policy for combating cyclical fluctuations, there are a number of things that can be done, both to bring about greater stability and achieve "prosperity without inflation and without war," to repeat some words which might well be taken as more than rhetoric. For one thing, cut massive military expenditures, which have peculiar inflationary properties. In addition to adding to the income-expenditure stream without offering corresponding supplies of goods to private markets, the nature and aura of military purchases and the symbiotic relation of public buyers and private suppliers offer a strong suggestion of inflationary potential. But government taxes which relate directly to final products or factors clearly raise prices. One could stimulate demand *and* combat inflation by cutting excise taxes and payroll taxes and by subsidizing certain production, such as mass transit services, which may be particularly desirable in terms of positive externalities (or the ability to substitute for other production with negative externalities). Further, we could abandon the curious combination of high interest rates and business taxes, which in the long run must in at least considerable part be passed on to consumers, and special tax subsidies, such as equipment tax credits and accelerated depreciation, which compensate or more than compensate for some these charges which are imposed on all. And finally, while I have not introduced distributional considerations, I might venture a considered judgment that the package of policies to which I have pointed would on balance cater well to egalitarian tastes.

If economists can free themselves from old *political* ideas and advocate boldly what makes economic sense, the Keynesian revolution, appropriately refined and revised, may yet carry the day. For "soon or late, it is ideas, not vested interests, which are dangerous for good or evil."

REFERENCES

R. W. Clower, "Is There an Optimal Money Supply?," *J. of Fin.*, May 1970, *25*, 425–33.

E. D. Domar, "Capital Expansion, Rate of Growth, and Employment," *Econometrica*, Apr. 1946, *14*, 137–47.

R. Eisner, "Another Look at Liquidity Preference," *Econometrica*, July 1963, *31*, 531–38.

———, "On Non-linear Estimates of the Liquidity Trap," *Econometrica*, Sept. 1971, *39*, 861–64.

M. Friedman, *A Theory of the Consumption Function*, Princeton 1957.

———, "The Role of Monetary Policy," *Amer. Econ. Rev.*, Mar. 1968, *58*, 1–17.

J. K. Galbraith, *The New Industrial State*, 2d ed., Boston 1971.

J. M. Keynes, *The General Theory of Employment, Interest and Money*, New York 1936.

D. W. Jorgenson, "Capital Theory and Investment Behavior," *Amer. Econ. Rev. Proc.*, May 1963, *53*, 247–59.

A. S. B. Leijonhufvud, *On Keynesian Economics and the Economics of Keynes*, New York 1968.

F. Modigliani and R. Brumberg, "Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data," in K. K. Kurihara, ed., *Post-Keynesian Economics*, New Brunswick 1954, 388–436.

D. Patinkin, *Money, Interest and Prices*, 2d ed., New York 1965.

R. M. Solow, "A Contribution to the Theory of Economic Growth," *Quart. J. of Econ.*, Feb. 1956, *70*, 65–94.

J. Tobin, "A Dynamic Aggregative Model," *J. Polit. Econ.*, Apr. 1955, *63*, 103–15.

Keynesian Models of Recession and Depression

By JAMES TOBIN*

Keynes's *General Theory* attempted to prove the existence of equilibrium with involuntary unemployment, and this pretension touched off a long theoretical controversy. A. C. Pigou, in particular, argued effectively that there could not be a long-run equilibrium with excess supply of labor. The predominant verdict of history is that, as a matter of pure theory, Keynes failed to prove his case.

Very likely Keynes chose the wrong battleground. Equilibrium analysis and comparative statics were the tools to which he naturally turned to express his ideas, but they were probably not the best tools for his purpose. For one thing, he explicitly confined the *General Theory* to a time period in which are given "the existing skill and quantity of available labor, quality and quantity of available equipment, the existing technique" and other factors. As he said (p. 245), "in this place and context, we are not considering or taking into account the effects and consequences of changes in them." But his model produces a solution in which, in general, the stock of capital, and other stocks, are not constant. Changes in these stocks will in turn alter investment, saving, and other behavior. For this reason alone, the solution of Keynes's model cannot be stationary, even in its own endogenous variables; and on this ground alone, it fails to qualify as an equilibrium. The evolution of Keynesian equilibrium as stocks change is receiving a great deal of attention these days and I shall not dwell on this point here. (See, however, A. S. Blinder and R. M. Solow and J. Tobin and W. Buiter.)

The second important point, the one on which Pigou insisted, is that excess supply of labor must cause money wages to decline. Even if this did not succeed in eliminating unemployment, one might not call a situation in which money wages and prices are persistently falling an equilibrium. But of course Pigou went further in contesting Keynes's claim that a "trap" might exist from which the economy could not be rescued, however low the wage and price level.

Keynes tried to make a double argument about wage reduction and employment. One was that wage rates were very slow to decline in the face of excess supply. The other was that, even if they declined faster, employment would not—in depression circumstances—increase. As to the second point, he was well aware of the dynamic argument that *declining* money wage rates are unfavorable to aggregate demand.[1] But perhaps he did not insist upon it strongly enough, for the subsequent theoretical argument focused on the statics of alternative stable wage levels.

The real issue is not the existence of a

---

* Sterling Professor of Economics, Yale University. The research described in this paper was undertaken by grants from the National Science Foundation and the Ford Foundation.

[1] ". . . it would be much better that wages should be rigidly fixed and deemed incapable of material changes, than that depression should be accompanied by a gradual downward tendency of money-wages, a further moderate wage reduction being expected to signalise each increase of, say, 1 percent in the amount of unemployment. For example, the effect of an expectation that wages are going to sag by, say, 2 percent in the coming year will be roughly equivalent to the effect of a rise of 2 percent in the amount of interest payable for the same period. The same observations apply *mutatis mutandis* to the case of a boom." (See Keynes, p. 265.)

long-run static *equilibrium* with unemployment, but the possibility of protracted unemployment which the natural adjustments of a market economy remedy very slowly if at all. So what if, within the recherché rules of the contest, Keynes failed to establish an "underemployment equilibrium"? The phenomena he described are better regarded as disequilibrium dynamics. Keynes's comparative statics were an awkward analytical language unequal to the shrewd observations and intuitions he was trying to embody. If the purity of neoclassical equilibrium is preserved, this verdict is no real blow to Keynes or solace for Pigou. The Great Depression is the Great Depression, the notorious "Treasury View" is still ridiculous, whether mass unemployment is a feature of an equilibrium or of a prolonged disequilibrium.

The issue is by no means dead. Today "full employment" has become the "natural rate," and "equilibrium" often allows for any steady rate of deflation or inflation, not just zero. But the proposition which Keynes was questioning is once again strongly argued in the profession and in public debate. Once again it is alleged that the private market economy can and will, without aid from government policy, steer itself to full employment equilibrium. This is the basis for advocacy of fixed rules of monetary growth and fiscal policy, as against active discretionary policy responding to information fed back from the private economy. At this very moment it is the basis for a policy of letting the recession run its course, in confidence that in a relatively short run—two or three years—equilibrium will be restored at full employment with reduced or even zero inflation.

I. Keynesian and Marshallian Price Dynamics

Milton Friedman (p. 18) has pointed out that Keynes was a "Marshallian in method" and translated the supply-demand framework of Alfred Marshall from individual markets to the whole economy. "Where he deviated from Marshall, and it was a momentous deviation, was in reversing the roles assigned to price and quantity. He assumed that, at least for changes in aggregate demand, quantity was the variable that adjusted rapidly, while price was the variable that adjusted slowly, at least in a downward direction." Friedman is correct that this was a momentous deviation, and one way to appreciate the point is to look explicitly at the dynamic stability implications of Walrasian vs. Marshallian assumptions about quantity adjustment.

Marshallian adjustment in a particular market is that quantity adjusts to the difference between demand price and supply price for existing quantity. Walrasian adjustment is that quantity adjusts to the difference between demand and supply at existing price.

Let us now apply these two adjustment assumptions to a simple macroeconomic model. Let Y be aggregate real output, and Y^* its value at full employment, i.e., at the "natural rate" level of unemployment. Let E be aggregate real effective demand, which can differ in short-run disequilibrium both from Y and from Y^*. Given the nominal stock of outside money M and other exogenous or policy-set variables, effective demand E is a function $E(p, x, Y)$ of three variables: p the price level, x its expected rate of change, and Y the level of output and real income.

In finer detail, E is the sum of consumption C, private investment I, and government purchases G:

$$(1) \quad E = C\left(Y, Y^*, -T, -R, x\frac{M}{p}, W\right) + I(Y, Y^*, -K, -R) + G$$

Here the C and I functions have positive derivatives in all their arguments. T represents taxes, a function of Y and Y^*. W is private wealth, equal to

$$\frac{M}{p} + qK,$$

where the coefficient q is the ratio of market valuation of capital equity to replacement cost. An increase in the real interest rate R relative to the marginal efficiency of capital makes q fall, and makes investment fall. The marginal efficiency of capital depends positively on Y and Y^*, negatively on K. The real interest rate R depends inversely on both M/p and x, and rises with Y and W.

The *price level effect* E_p on demand is negative, for the following familiar combination of reasons. First is the Keynes effect. A given nominal quantity of money will be a larger real quantity at a lower price level. Consequently the interest rate may be lower, and investment demand higher. The Keynes effect is expected to be weaker the larger the real supply of money relative to output Y, and to vanish altogether in the "liquidity trap." This will tend to make E_p smaller in absolute value at low levels of Yp/M.

Second is the Pigou effect, the wealth effect on consumption. The lower the price level, the higher the real value of those components of net private wealth fixed in nominal value. The relevant components are outside money (and some part of any nonmonetary public debt in existence). Consumption demand is expected, *ceteris paribus*, to respond positively to increases of wealth.

The short-run Pigou effect is very likely weaker than the long-run effect and may not even have the same negative sign. And it is the short-run effect which is relevant for Keynesian theory and for the dynamics of this paper. The difference arises as follows: among the stocks fixed in the short run are private debts in the unit of account. These are a heavier burden to debtors the lower the price level, and there are good reasons why transfer of real income and wealth to creditors spells a net deficit of aggregate demand. Debtors are debtors because they have high propensities to spend. Many of them are liquidity-constrained, and as their debt/equity ratios increase their credit lines dwindle or, in case of bankruptcies, disappear. Although these are "only" distributional effects, they may be more important than the real value of outside money and debt.

The long-run comparative-static Pigou effect, in contrast, assumes that each alternative price level has prevailed for a sufficiently long time so that inside debts are scaled to that price level—although strangely enough exogenous outside money is not. In this counterhistorical "as if" mental experiment, debtors are no more burdened at one price level than at another.

As for *the price change effect* E_x, there are several effects. A decrease in the expected inflation rate raises the real rate of interest. This increase discourages investment, and it also deters consumption both directly and by lowering the market value of equity capital, one component of wealth. On the other hand, expected capital gains on money holdings xM/p are favorable to consumption. This is a "flow Pigou effect," to be distinguished from the stock effect. The question here involves the size of the marginal propensity to spend from expected real capital gains. Econometric evidence has been that this marginal propensity is small, although capital gains eventually affect consumption via the wealth effect. I have assumed that the other effects of expected inflation dominate the flow Pigou effect.

The *marginal propensity to spend* E_y is taken to lie between 0 and 1 on usual Keynesian grounds. As is well-known, a high response of investment demand to contemporaneous income could easily make E_y exceed one. But Keynes typically regarded investment as determined more by long-run sales and profit expectations than by current business activity. The likelihood that, in prolonged departures from full employment, investment will come to be governed more by contemporaneous than by full employment sales and profits is a source of possible instability and of prolonged disequilibrium to which I shall return later in the paper.

In equilibrium, the following three conditions hold:

$$E(p, x, Y) - Y = 0 \tag{2.1}$$

$$Y - Y^* = 0 \tag{2.2}$$

$$x = \dot{p}/p = 0. \tag{2.3}$$

(I shall also denote $\dot{p}/p$ as π.)

I shall call the first dynamic version of this model the *WKP* model (Walras-Keynes-Phillips). All the adjustment functions which follow will conform to the notation $A_y z$, where y is the variable adjusting, z the variable on which the adjustment depends, and A_y a positive constant. The *WKP* model is as follows:

The WKP Model

$$\dot{Y} = A_y(E - Y) \tag{2.1.1}$$

$$\pi = A_p(Y - Y^*) + x \tag{2.2.1}$$

$$\dot{x} = A_x(\pi - x) \tag{2.3.1}$$

Equation (2.1.1) says that production Y moves in response to discrepancies of E and Y. This implements the Keynesian view that in the very short run money wages and prices are set and output responds to variations of demand.

How can E and Y diverge even for an instant? Many words have been spilled, both by Keynes himself and by others, on this question, usually posed in terms of the possibility of inequality of Saving and Investment. In our present context, let D be the demand which must always equal Y to preserve the national income identities. Let D be a function of $\dot{Y}$ as well as of x, p, and Y. Then $D(\dot{Y}, x, p, Y) = Y$, $E(x, p, Y) = D(0, x, p, Y)$. Equation (2.1.1) follows from a negative value of $\partial D/\partial \dot{Y}$, which means that demand is lower, at given Y, when Y is increasing. Lags in consumption spending lead to this sign and so does *unintended* inventory decumulation. The investment accelerator works in the other direction, but for the reason already given it is not a Keynesian idea.

Equation (2.2.1) is a natural-rate version of the Phillips curve. The short-run Phillips curve is the obvious Keynesian version of price dynamics. Throughout this paper I am condensing product and labor markets into one sector and assuming with Keynes that prices are determined by marginal variable costs, i.e., by labor costs. Excess labor supply and $Y - Y^*$, the "Okun gap," are linked,—when one is zero so is the other. So it is the gap which causes wage rates to fall. But to "fall" does not mean to decline absolutely; it means to decline relative to x, the accustomed and expected rate of inflation of both labor costs and prices. This is the more modern wrinkle. By here assuming (2.2.1) I do not mean necessarily to associate myself—much less Keynes!—with the natural-rate hypothesis in all its power and glory.

The third equation (2.3.1) is the well-known model of adaptive expectations. There is nothing particularly Keynesian about this equation, and the same formulation will carry over to the non-Keynesian dynamic model. Keynes himself would scorn it and stress instead the stochastic and historical sources of expectations. But like so many of his observations, these do not lend themselves to simple formal analysis.

As two extremes of interest I shall wish to consider:

$$x = \pi \tag{2.3.2}$$

(extrapolation of current rates of price *change*)

$$x = 0 \tag{2.3.3}$$

(extrapolation of current price *level*)

The alternative dynamic version may be called the M model (Marshall). The equations are:

The M Model

$$\pi = B_p(E - Y) + x \tag{2.1.2}$$

$$\dot{Y} = B_Y(Y^* - Y) \tag{2.2.2}$$

$$\dot{x} = A_x(\pi - x) \quad \text{(or 2.3.1 or 2.3.3)} \tag{2.3.2}$$

As compared with the WKP model, the adjustment roles of the first two equations are interchanged. The first equation now says that the immediate impact of excess demand for goods and services is to raise prices, or more strictly to raise them faster than they had been expected to rise. (It is not entirely accurate to regard (2.1.2) as non-Keynesian. When there is an inflationary gap ($E>Y^*$, $Y=Y^*$), this looks very much like the Keynesian model of inflation. But in Keynes's inflation theory, Y^* is considered an absolute short-run constraint on production, as in wartime. In normal conditions, Keynes would, I think, regard Y^* as a medium-run labor market equilibrium with normal margins of excess capacity and of frictional unemployment, a level of output which could be at least temporarily exceeded.)[2] In any event, equation (2.1.2) is one way to inject into the model the view that prices respond quite flexibly to changes in excess demand for goods, whether or not the economy is close to full employment.

The non-Keynesian partner of this price adjustment equation is (2.2.2), where the gap between potential and actual output inspires adjustments of production and employment. This is because they are associated with gaps of the same sign between the demand price for labor (the value of its marginal product) and its supply price.[3] The idea is that when Y^* exceeds Y the real wage is less than marginal productivity. Competitive employers therefore add to their work forces and their production. In Keynesian theory, on the other hand, production increases only when demand at existing prices expands.

II. Local Stability of the Two Models

Let us now consider the local stability of the WKP and M models, around their equilibrium values $Y=Y^*$, $p=p^*$, $x=0$. For this purpose it is convenient to substitute in the third equation the value of $\pi - x$ drawn from the second or first equation. Thus the third equations in the WKP and M models become respectively:

$$\dot{x} = A_x A_p(Y - Y^*) \tag{3.1}$$

$$\dot{x} = A_x B_p(E - Y) \tag{3.2}$$

For the WKP model, the linearized equations are:

$$\begin{bmatrix} \dot{Y} \\ \dot{p} \\ \dot{x} \end{bmatrix} = \begin{bmatrix} A_Y(E_y - 1) & A_Y E_p & A_Y E_x \\ A_p p^* & 0 & p^* \\ A_x A_p & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} Y - Y^* \\ p - p^* \\ x \end{bmatrix} \tag{3.3}$$

The critical necessary condition for stability is:

$$p^* E_p + A_x E_x < 0 \tag{3.4}$$

The first term of (3.4) is negative and the second term positive. As would be expected, a strong negative price-level effect on aggregate demand, a weak price-expectation effect, and a slow response of price expectations to experience are conducive to stability. In one extreme case (2.3.3), where $x=\dot{x}=0$, the system is of course stable. In the other extreme case (2.3.2), where $x=\pi$, the first term of (3.4) drops out and the system is necessarily unstable.

---

[2] In the *General Theory*, Keynes discusses frictional and involuntary unemployment on p. 6 and, in defining involuntary unemployment on p. 15, says, "Clearly we do not mean by 'involuntary' unemployment the mere existence of an unexhausted capacity to work."

[3] This is true even if the labor supply curve is downward sloping, provided it is closer to vertical than the schedule of marginal productivity of labor.

The M model is quite different. It is separable into output and price equations. Equation (2.2.2) is a stable differential equation in the single variable Y. The stability of the price system depends on (3.4), in the same way as the stability of the WKP model. The formal system is:

$$(3.5)\quad \begin{bmatrix} \dot{Y} \\ \dot{p} \\ \dot{x} \end{bmatrix} = \begin{bmatrix} -B_Y & 0 & 0 \\ B_p(E_Y-1)p^* & B_pp^*E_p & B_pp^*E_x+p^* \\ A_xB_p(E_y-1) & A_xB_pE_p & A_xB_pE_x \end{bmatrix} \cdot \begin{bmatrix} Y-Y^* \\ p-p^* \\ x \end{bmatrix}$$

As Friedman surmised, Keynes's choice of adjustment mechanisms is a crucial element of his theory. In particular, the Walras-Keynes-Phillips adjustment model allows the distinct possibility that lapses from full employment will not be automatically remedied by market forces. Keynes could also be interpreted to hold the view that price-level effects E_p are weak relative to speculative effects E_x. I shall discuss this interpretation further in the next section.

III. Irreversible Recessions and Deep Depressions

Let us take a more global look at the equilibrium condition $E=Y$ (2.1). In Figure 1 are shown in (p, x) space several loci along which the condition is met. The slope of such a locus, $-E_x/E_p$, is positive. Each locus is for a given value of Y; a reduction in Y shifts the locus to the left. In the Figure, the right-most locus is for full employment output Y^*. The weakening or vanishing of the "Keynes effect" at low values of Y and p tends to reduce E_p in absolute value. This is reflected in the curvature of the loci.

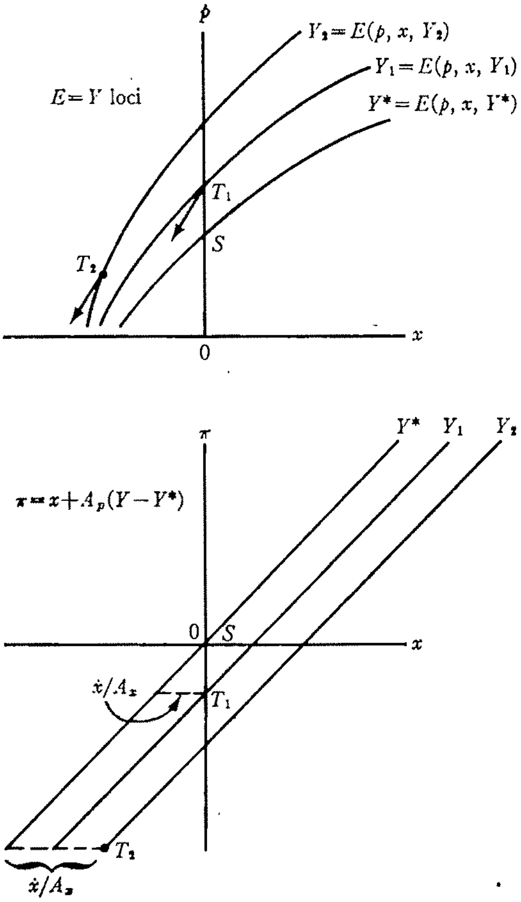


FIGURE 1

Consider an initial position T_1 at levels of E and Y short of Y^*. Prices begin to decline because Y_1 is less than Y^*. To a degree that depends on the speed of adaptation, expectations of price change become negative. The arrow indicates the direction of movement. As drawn, the movement is stabilizing, taking the economy to higher E and Y, toward Y^* and the equilibrium S.

The lower panel of Figure 1 concerns the direction of the arrow, the relationship of π and x. The horizontal axis matches in origin and scale that of the upper panel. The lines are parallel 45 degree lines, for Y^*, Y_1, and Y_2, the same output levels as in the upper diagram. The points S, T_1, T_2 correspond to the similarly labeled points above. At S, $\pi=0$. At T_1, π is negative. So is $\dot{x}$, by an amount proportional to the difference between π and x, shown horizontally as $\dot{x}/A_x$.

Consider instead an initial position T_2 in the two panels. At T_2 both the slope of the $E=Y$ locus and that of the arrow are steeper. The reason that the arrow is steeper can be seen in the lower panel: $\dot{x}/A_x$ has doubled, but π has more than doubled. The net outcome could go either way. The possibility illustrated is that at T_2 the locus $E=Y$ is so steep that the movement is destabilizing. The system might be stable for small deviations from its equilibrium but unstable for large shocks.[4] The failure of automatic market processes to restore full employment would be reinforced if large and prolonged recession caused investors to gear their estimates of the marginal efficiency of capital more to current than to equilibrium demand and profitability.

Under these adverse circumstances, and in the absence of countercyclical policy, the economy could slip into a deep depression.

In nonlinear nonmonetary business cycle models like those of M. Kalecki, R. Goodwin, and Sir John Hicks, a long depression phase occurs with the economy at a floor. At this floor the capital stock is excessive and gross investment is zero; production is solely to meet minimal private and social consumption requirements, which are independent of income and wealth. The depression phase lasts a long time, while depreciation slowly whittles the capital stock down to the amount needed for floor level production.

It is not part of this paper to provide a model of such a floor. The relevant question is whether deflation will by itself lift the economy from the floor. Will deflation so augment private wealth that consumption rises above its floor level? Clearly this will not happen unless condition (3.4) is met at the depression income level.

But at the floor, E_x is higher than in the normal regime. An increase in the deflation rate $-x$ lowers the value of the capital stock. The physical capital stock declines slowly. But its value—its *real* value—can decline rapidly; when no gross investment is taking place, the existing stock will be valued well below replacement cost. At the liquidity trap, the real interest rate is the irreducible nominal rate $\bar{r}$ plus the expected rate of deflation $-x$. The value of a unit of capital is $(\rho-\delta/(r-x))$ where $\rho-\delta$ is the marginal productivity of capital net of depreciation. Although the attrition of the stock slowly raises ρ, deflation rapidly raises $\bar{r}-x$.

IV. Concluding Remarks

God may have made the world so that full employment equilibrium exists and is stable. Perhaps the divine design guarantees that capitalist market economies will never be trapped in depressions with involuntary unemployment and will never need to depart from fixed no-feedback rules of fiscal and monetary policy. But Keynes had good empirical and theoretical reason to suspect otherwise. He did not establish an underemployment equilibrium. But he did not really need to. Even with stable monetary and fiscal policy, combined with price and wage flexibility, the adjustment mechanisms of the economy may be too weak to eliminate persistent unemployment.

---

[4] Robert Solow has pointed out to me that the possibility illustrated by T_2 is only suggestive of a global instability. The global properties of the system require further investigation.

REFERENCES

A. S. Blinder and R. M. Solow, "Does Fiscal Policy Matter?," in *J. of Pub. Econ.*, 1973, *2*, 319–37, and *Econ. of Pub. Fin.*, Brookings Institution, 1974, 452–58.

M. Friedman, *A Theoretical Framework for Monetary Analysis*, Nat. Bur. of Econ. Res. occas. pap. 112, New York 1971.

J. M. Keynes, *The General Theory of Employment, Interest, and Money*, 1st ed., London 1936, 245.

J. Tobin and W. Buiter, "Long-Run Effects of Fiscal and Monetary Policy on Aggregate Demand," Cowles Foundation disc. pap. no. 384, Dec. 1974.

PERSONAL SAVINGS IN THE POSTWAR WORLD: IMPLICATIONS FOR THE THEORY OF HOUSEHOLD BEHAVIOR

Towards a Theory of Saving Behavior

By F. THOMAS JUSTER AND LESTER D. TAYLOR*

The behavior of personal saving in the *U.S.* economy has long held a fascination for economists. The earliest consumption functions, based on cross-sectional data from household budget surveys, indicated that richer households save more relative to income than poorer ones, leading to the Keynesian speculation that secularly rising income would result in higher saving rates. The monumental study of *U.S.* capital formation of Simon Kuznets indicated that this attractive theory did not square with the facts, since there was no evidence that saving as a fraction of income increased as income rose secularly.

This apparent contradiction between cross-sectional and long-term time-series evidence attracted a great deal of attention in the late 1940's and early 1950's and led to the formulation of three new theories of the consumption function, the relative income hypothesis of J. S. Duesenberry, the life-cycle hypothesis of F. Modigliani and R. Brumberg, and the permanent-income hypothesis of Milton Friedman.[1] Although these theories, especially the last two, have inspired a huge literature, its main strands really have had little to say about cyclical movements in personal saving, as opposed to secular movements.[2]

While focusing on the secular behavior of saving was understandable in the 1950's and early 1960's when the personal saving rate moved within fairly narrow limits, the saving rate in recent years has been considerably more volatile. This cyclical volatility has been a concern of both Juster and Taylor in recent years, and the present undertaking represents a merging of efforts. The focus of Taylor's work has been on the role of income composition in determining the behavior of aggregate saving, while Juster has concentrated on the role of uncertainty.[3] The present paper attempts to combine both of these insights into a more refined model of personal saving.

I. Income Composition and Adjustment Effects

Excluding the effects of uncertainty, which will be discussed below, the general framework for our analysis is the "zero-depreciation" theory of saving elaborated by H. S. Houthakker and Taylor. According to this theory, personal saving (S) is assumed to be a linear function of the existing stock of financial assets (A) and income (Y):

$$S(t) = \alpha + \beta A(t) + \gamma Y(t). \tag{1}$$

The stock of financial assets is assumed to be nondepreciating, with the result that the relationship between the stock A and its flow S is given by

---

* University of Michigan and University of Arizona, respectively.

[1] A fourth theory, developed by James Tobin, involves an extension of the absolute-income hypothesis of Keynes that takes into account the household's level of wealth.

[2] For a detailed review of the permanent-income hypothesis and critical evaluation of the many and various attempts to test it, see T. Mayer.

[3] See Taylor (1971 and 1974). Also, see Juster (1973) and Juster and P. Wachtel (1972a and 1972b).

$$(2) \qquad \dot{A}(t) = S(t).$$

The basic notion underlying the model is that the consumer adjusts saving to bring the stock of financial assets into line with income. This model can be interpreted as embodying a Stone-Nerlove stock adjustment process to the stock of wealth, with the proviso that the depreciation rate on the stock is zero. An important consequence of this assumption is that saving is zero in long-run, steady-state equilibrium. Accordingly, the concept of a marginal propensity to save must be with reference to some limited interval of time. More particularly, γ in equation (1) refers to the instantaneous marginal propensity to save defined by the condition that the underlying stock of financial assets is constant. However, γ should not be confused with the one-quarter marginal propensity to save which will be the focus in the empirical work.

Following Taylor (1971), one of the objectives of our inquiry requires disaggregation of disposable income into the following components: labor income (L), property income (PR), transfers (TR), personal contributions to social insurance (SI), and personal tax and nontax payments (T). With this decomposition of disposable income, equation (1) becomes

$$(3) \quad S(t) = \alpha + \beta A(t) + \gamma_1 L(t) + \gamma_2 PR(t) + \gamma_3 TR(t) + \gamma_4 SI(t) + \gamma_5 T(t).$$

Since savings are the epitome of a durable good, β should be negative: the stock of wealth should have the same (qualitative) effect on saving that the stock of automobiles, say, has on the sale of new cars. The coefficients γ_1, γ_2, and γ_3 should, of course, be positive and on theoretical grounds γ_2 should be greater than γ_1 because property income is more uncertain than labor income. The marginal propensity to save out of transfer payments has traditionally been thought to be low, but all empirical work to date with this model has found γ_3 to be in the neighborhood of .8. Finally, γ_4 and γ_5 are naturally expected to be negative, and since contributions to social insurance leave the contributors with a claim to future income—rather than just a tax receipt—the size of γ_4 will depend upon the extent to which social insurance substitutes for regular saving. At an extreme, γ_4 could be in the neighborhood of -2, since employees' contributions are matched dollar-for-dollar by employers.

As formulated in (3), the model contains a quantity, the stock of financial assets, that can at best be measured with difficulty, hence it is desirable that it be eliminated. Once this is done, and also after the model is translated into one involving discrete intervals of time, the estimating form of the model, with an error term (u) added, becomes:[4]

$$(4) \quad S_t = B_1 S_{t-1} + B_2 \Delta L_t + B_3 \Delta PR_t + B_4 \Delta TR_t + B_5 \Delta SI_t + B_6 \Delta T_t + u_t.$$

The coefficients $B_1, \ldots, B_6$ are functions of the structural parameters $\beta, \gamma_1, \ldots, \gamma_4$ (α is indeterminate), and estimates of the latter can be readily obtained from estimates of the former.

II. The Role of Uncertainty

Economists have long paid lip service to the role of uncertainty in affecting economic behavior, but attempts to quantify the impact of uncertainty have been few and far between. Empirically, uncertainty regarding income usually takes the form of supposing that people make judgments about the future on the basis of past behavior of the quantities in question. However, for reasons that need not be discussed at length, the customary procedures seem to us to be seriously deficient, and we feel that there may be some real payoff in devising direct measures of uncertainty that are both relevant and operational.

---

[4] The details of this elimination and translation are set out in Houthakker and Taylor, pp. 13–17. For the formulae involved, see p. 17.

Following Juster and P. Wachtel (1972a) and Juster (1973), our approach takes into account two types of uncertainty that are expected to influence saving behavior. The first type is the conventional kind of uncertainty regarding income. If consumers become concerned that money income may fall, they can be expected to react by trying to build up cash balances via increased saving. If money income actually falls, the fixed-cost elements of consumption expenditures can be expected to result in lower saving. Empirical representation of these effects can be accomplished by using the unemployment rate to measure money income that is high or low relative to the recent past, and to use the change in unemployment rate as a measure of changes in the degree to which consumers attach uncertainty to the receipt of future money income. Thus, unemployment is expected to have a negative impact on current saving, since it represents a negative transitory income effect, while the change in unemployment is expected to have a positive impact on saving.

The second, less conventional, and perhaps more important, dimension of uncertainty has to do with the effects of price inflation on saving. A recent paper by Wachtel (1974) suggests a useful classification for the effects of price inflation on saving: a money-illusion effect, an intertemporal substitution effect, and an uncertainty effect.

The first two of these effects are traditional subjects of analysis, but the third is a relatively recent and more speculative insight and is based on an idea long espoused by George Katona that rising pessimism regarding the future results in higher rates of saving. As elaborated by Juster (1973), the key ingredient is the downward inflexibility of money wage rates. In a world where money wage rates cannot fall, a zero rate of price inflation implies that changes in real wage rates cannot be negative. Consequently, uncertainty about the change in real income is limited to uncertainty about how large an increase in money wages will be forthcoming.

However, when prices are rising, real income can fall if money wages fail to rise as rapidly. Moreover, the higher the rate of price inflation, the greater the possible downward movement in real income. Viewed more generally, the distribution of the expected change in real wages is determined jointly by the distribution of the expected change in money wages, the distribution of the expected change in prices, and the relationship between these two quantities. Only in the extreme case of automatic and complete indexing will the existence of price inflation (of uncertain magnitude) not lead to greater uncertainty about real wages than in a world where price inflation is absent.

While price inflation per se is sufficient to generate uncertainty about the change in real income, the measure that best represents the kind of uncertainty that impacts on saving behavior might well be the maximum price rise considered "plausible" by consumers. That is, saving may be more influenced by judgments about plausible upper limits of future price increases than by estimates of the most probable change in prices. Although we do not have a direct measure of these upper limits, it does not seem unreasonable to suppose that the variation among households in the expected change in prices provides some useful indirect information.[5]

III. Empirical Results

The empirical analysis employs quarterly data, mostly taken from the National Income Accounts (*NIA*), and covers the period beginning with the second quarter of 1953 and ending with the fourth quarter of 1973. The *NIA* data are aggregates, seasonably adjusted at annual rates, expressed in billions of current dollars. The price expectations data come from surveys conducted periodically by the Survey Research Center at the University of Michigan.[6] The conventional *NIA* definition of personal saving is the saving concept that is analyzed.

Of the numerous equations estimated, five have been selected for presentation and discussion and are tabulated in Table 1. The first equation is included for purposes of comparison with Taylor's earlier efforts with the model based on equation (3) above. The relevant equation from Taylor (1971) is as follows: (t-ratios are in parentheses):

$$\begin{aligned}(4)\quad S_t = \underset{(43.79)}{.955S_{t-1}} &+ \underset{(4.21)}{.449\Delta L_t} + \underset{(0.86)}{.277\Delta PR_t} \\ &+ \underset{(2.86)}{.893\Delta TR_t} - \underset{(-3.30)}{2.16\Delta SI_t} \\ &- \underset{(-4.87)}{.901\Delta T_t} + \underset{(2.03)}{3.65\Delta BAA_t} \\ &R^2 = .899 \quad S_e = 2.01 \quad df = 61.\end{aligned}$$

The period of fit for this equation was 1953–I through 1969–IV.

The most notable difference between the first equation in Table 1 and equation (4) involves the coefficients for labor income and property income. The coefficient on ΔPR is now greater than the coefficient on ΔL, which is what we expect a priori, and now has a t-ratio approaching 2. This emergence of property income as a somewhat energetic predictor would seem to result from recent sharp increases in farm proprietors' income. Apart from this, the coefficient on the change in payroll taxes increases, but nevertheless remains less than -1. Note, too, that the impact of a change in transfers on saving is still substantial.

The remaining four equations in Table 1 incorporate the various uncertainty measures just discussed. Equations were estimated with ΔU as a predictor as well as U, but it was invariably weak statistically and therefore has been excluded

---

[5] For example, if all households expect prices to rise at the rate of zero percent per year, this may not be very different in terms of the impact on saving than if, *ceteris paribus*, all households expect prices to rise at a rate of 10 percent per year. But if there is a great deal of dispersion in "best guesses" about the extent of future price rises, this dispersion may be taken as a reasonable proxy for concerns about plausible upper limits. In any event, this variation across households represents the only empirical measure of uncertainty that we have, except, of course, for the mean expected price change across all households.

[6] The questions are relatively straightforward: "Do you expect prices to rise or fall or what?" and, starting with the first quarter of 1966, "How much do you expect prices to rise—1 or 2 percent, 3 or 4 percent, 5 percent, between 5 and 10 percent or more?" Estimates of the mean and variance of expected price change among the population sampled can be constructed either by assignment of quantitative magnitudes to basically qualitative responses for part of the period (i.e., until 1966) and assignment of midpoints for that part of the period for which quantitative data are available or by a less arbitrary, but perhaps also less satisfactory, procedure of throwing away all quantitative detail and focusing simply on responses in the up, down, no-change categories. The latter, concentrating only on direction of change, can be converted into estimates of means and variances, provided that one is willing to assume a specific distribution for expected price within the population sampled.

Both types of estimates have been used in the empirical analysis. As it turns out, the best results are obtained with the mean constructed by the second procedure and the variance constructed according to the first. The question of which procedure to employ is far from closed, however, and both we and others are attempting to improve upon our methods for constructing estimates of expected price change from the available survey information.

TABLE 1—REGRESSION RESULTS OF U.S. PERSONAL SAVING
Quarterly 1953-II—1973-IV
(t-ratios in parentheses)

Independent Variable	Equation				
	1	2	3	4	5
S_{t-1}	0.978	0.817	0.887	0.831	0.889
	(62.60)	(13.98)	(18.87)	(14.29)	(19.06)
ΔL_t	0.245	0.389	0.430	0.316	0.409
	(2.61)	(4.37)	(4.91)	(3.89)	(4.11)
ΔPR_t	0.400	0.266	0.119	0.469	0.385
	(1.99)	(1.22)	(0.53)	(2.23)	(1.76)
ΔTR_t	0.774	0.757	0.861	0.786	0.865
	(4.06)	(4.53)	(5.13)	(4.74)	(5.19)
ΔSI_t	−1.38	−0.791	−0.740	−0.874	−0.833
	(−2.89)	(−1.89)	(−1.75)	(−2.13)	(−2.00)
ΔT_t	−0.784	−0.773	−0.808	−0.752	−0.772
	(−5.43)	(−6.23)	(−6.53)	(−6.15)	(−6.25)
ΔBAA_t	3.30	2.67		3.57	
	(2.38)	(1.91)		(2.82)	
BAA_t		−0.738		−0.959	
		(−1.99)		(−2.42)	
$\Delta BAA_t Y_t$			0.0064		0.0082
			(2.15)		(3.14)
$BAA_t Y_t$			−0.0022		−0.0025
			(−3.43)		(−3.62)
ΔCPI_t^*		170.97			
		(1.74)			
$\Delta CPI_t^* Y_t$			0.495		
			(1.98)		
PE_t				0.492	
				(2.25)	
$PE_t Y_t$					0.0012
					(2.22)
VPE_t		0.840		0.969	
		(4.57)		(5.69)	
$VPE_t Y_t$			0.0011		0.0015
			(3.15)		(5.20)
U_t		0.501		0.510	
		(2.61)		(2.78)	
$U_t Y_t$			0.0016		0.0014
			(2.85)		(2.68)
R^2	0.953	0.969	0.969	0.970	0.970
S_e	2.33	1.96	1.95	1.93	1.94
df	76	72	72	72	72

Note: The variables are defined as follows: S = personal saving; L = labor income (equals sum of wage and salary disbursements and other labor income); PR = property income (equals proprietors' income, rental income of persons, dividends, and personal interest income); TR = transfer payments; SI = personal contributions to social insurance; T = personal tax and nontax payments; BAA = yield on Baa bonds; $CPI^* = (\Delta CPI_t + \Delta CPI_{t-1})/2$; PE = mean expected price change;[a] VPE = variance in mean expected price change within population sampled[a]; U = civilian unemployment rate.

[a] The methods of calculating PE and VPE are described in an appendix, obtainable on request from the authors.

from consideration. Equations 3 and 5 in the table have the yield on Baa bonds and the uncertainty measures interacted with the level of disposable income. The purpose of this was to allow for the possibility that the effects of uncertainty on saving vary systematically with the level of income. The results, however, do not suggest that this is the case. Finally, equations 2 and 3 differ from 4 and 5 in that 2 and 3 include the most recent six-month change in the consumer price index (*CPI*) as a predictor in place of the mean expected price change.

Addition of the uncertainty measures to the model leads to the following results and conclusions:

(1) Except for the unemployment rate, the signs of the uncertainty measures are all as expected a priori.[7] In particular, the results strongly support the conclusion that uncertainty regarding the future course of real and money income leads to an increase in saving.

(2) Of the uncertainty measures considered, the strongest statistically is the variance term. As suggested by the discussion in Section II, we interpret the strong performance of this variable as representing the impact on saving of uncertainty regarding plausible upper limits to the amount of price inflation.[8]

(3) The coefficient on payroll taxes is reduced considerably in absolute value and is now of the same magnitude as the coefficient on personal tax and nontax payments. Taking equation 4 as the point of reference, the only other coefficients to undergo much change (vis-à-vis equation 1) are those on S_{t-1} and labor income. The former is reduced to .83, which implies a speedier dynamic adjustment, while the latter is increased by about one-third. It is of particular interest that the inclusion of the uncertainty measures does not alter the statistical significance of the income terms. And it is especially to be noted that the coefficient on the change in transfer payments remains large.

(4) Finally, it is to be noticed that the coefficient for the yield on Baa bonds is negative, while the coefficient for its first difference is positive. Since the factors operating on the real rate of interest should be expected to change slowly through time, the first difference in the nominal rate of interest should, in view of Irving Fisher, be dominated by price expectations. This being the case, the positive coefficient on ΔBAA is very much in keeping with the views expressed in Section II. However, the negative coefficient for *BAA*, in an expectations framework, is harder to interpret.

REFERENCES

J. S. Duesenberry, *Income, Saving, and The Theory of Consumer Behavior*, Cambridge, Mass. 1949.

M. Friedman, *A Theory of the Consumption Function*, Princeton 1957.

H. S. Houthakker and L. D. Taylor, *Consumer Demand in the United States*, 2d ed., Cambridge, Mass. 1970.

F. T. Juster, "Savings Behavior, Uncertainty, and Price Expectations," *The Economic Outlook For 1974*, Univ. of Michigan 1973.

---

[7] The discussion in Section II viewed the unemployment rate as a proxy for (negative) transitory income, hence it was postulated that its sign should be negative. However, the fact that this is not the case suggests that the unemployment rate augments the price expectations variables as a direct measure of uncertainty.

[8] There are, of course, other hypotheses that can be advanced to account for the importance of *VPE*. In particular, we are especially interested in exploring the possibility that *VPE* represents an aggregation effect arising from variation in the impact of price expectations on saving across income groups. Nevertheless, the fact remains that, for whatever reason, *VPE* is an important predictor empirically.

——— and P. Wachtel, "Inflation and the Consumer," *Brookings Papers on Economic Activity,* No. 1, 1972a.

——— and ———, "A Note on Inflation and the Saving Rate," *Brookings Papers on Economic Activity,* No. 2, 1972b.

T. Mayer, *Permanent Income, Wealth, and Consumption,* Berkeley 1973.

F. Modigliani and R. Brumberg, "Utility Analysis and Aggregate Consumption Functions: An Attempted Integration," 1953.

L. D. Taylor, "Saving Out of Different Types of Income," *Brookings Papers on Economic Activity,* No. 2, 1971.

———, "Price Expectations and Households' Demand for Financial Assets," *Explorations in Economic Research,* Nat. Bur. of Econ. Res., No. 2, 1974.

J. Tobin, "Relative Income, Absolute Income, and Saving," in *Money, Trade, and Economic Growth: Essays in Honor of John H. Williams,* New York 1951.

P. Wachtel, "Survey Measures of Expectations Information and Their Potential Usefulness," paper presented at the Income and Wealth Conference, Nat. Bur. of Econ. Res., Washington, Nov. 1974.

Saving Behavior in Western Germany and the United States

By BURKHARD STRUMPEL*

Sharp cross-national differences in the saving behavior of private households are reasonably well documented but not well understood. The cross-country variance in the rate of saving has been explained only to a small extent by a country's income status, change, or trend. There are doubts if "one true world saving function in which saving is a function of current income" really exists (D. W. Johnson and J. S. Y. Chiu, p. 333).

Economic theory has repeatedly been criticized for the neglect of cultural and institutional constraints that restrict the validity of its laws. The problem of ethnocentricity has been discussed extensively in the fields of economic development and comparative economic systems, but much less so in other areas. The easy reaction to this criticism is to define time/space coordinates in order to limit the application of the theories in question. It is more constructive, however, to reformulate the theory in a fashion sufficiently general to incorporate the "deviant" cases. Starting out from the large and unexplained differences in the rates of household saving between Western industrial countries, this paper presents a model of household saving emphasizing its cultural and institutional determinants. It thus deals explicitly with some of the factors that shape a nation's saving behavior in reaction to given income conditions and that have not been treated by the well-known theories of the consumption function.

These theories have directed most effort towards alternative formulations of the explanatory variable income, its distribution across individuals in a society, its sources, and its change over time. Little attention has been given to the refinement of the dependent variables. Spending or saving, consumption or asset formation are mostly treated as aggregates only. They are not being differentiated according to their uses or their purposes as intended by the decision maker. According to Kevin Lancaster, utility is linked not only to a good but to the characteristics of a good as they are related to a variety of household goals or ends. Similarly, saving must be seen as instrumental for a variety of goals. Different people and nations, even those operating within similar income constraints, can be shown to differ widely with respect to the intensity of these goals, as well as the instrumental role of saving in reaching these goals. Why should the motivation to save for the purpose of acquiring a home respond to an income increase in the same manner as the inclination to accumulate reserves for emergencies or rainy days?

Adapting the expectancy-value approach, I see the motivation to save as dependent on:

(1) the strength of the goals that are to be served through saving (e.g., sense of security in case of emergencies; subsistence at old age; acquiring real estate or durables);

(2) the instrumentality of saving: To

---

* Program Director, Institute for Social Research, University of Michigan, and Visiting Professor of Economics, University of Frankfurt, Germany, for 1974–75.

what extent is saving perceived as indeed approaching or resulting in that goal?

(3) the strength of the motivation to spend; since saving and spending are subject to a common budget constraint, a strong motivation to spend will tend to depress saving, and vice versa.

This model will be used below to explain differences in rates of saving between Western Germany and the United States. The available evidence on goals and instrumentalities of saving, while suggestive, does not yet permit a quantitative test of the model as applied to a large number of countries or a long period of time. In this sense, the model is programmatic rather than definitive.

Table 1 exhibits some saving statistics in important Western countries and Japan for selected years between 1954 and 1973. As is well-known, there are considerable

TABLE 1—RATES OF POSTWAR SAVING IN FIVE WESTERN COUNTRIES[a]

(Figures given are percentage of aggregate disposable income)

	Gross financial saving (1)	Increase in consumer debt (2)	Net financial saving (3)	Total saving (4)	Social security contributions (5)	Aggregate[b] disposable income (annual average)
United States						Billions Dollars
1954–56	9.1	5.8	3.3	7.3		271
1964–66	9.6	5.7	3.9	6.2	6.9	464
1968–70	10.6	4.5	6.1	7.7	8.4	628
1973	14.8	7.8	7.0	6.2	10.4	882
Germany						Billions DM
1954–56	5.9	0.7	5.2	12.5		120
1966	9.8	0.3	9.5	14.6	15.2	326
1968–70	11.7	0.9	10.8	15.3	16.8	396
1973	12.3	0.7	11.6	14.1	21.2	578
United Kingdom						Millions Pounds
1954–56				1.8	4.4	13,208
1964	9.0	4.5	4.4	5.8	6.3	22,732
1968–70	9.2	3.7	5.5	5.3	7.7	30,464
1972	13.3	10.6	2.7	8.3	7.1	46,677
France						Billions Francs
1954–56				6.8		129
1964	9.0	2.2	6.7	10.6	19.8	311
1968–70	8.1	4.2	3.9	11.1	21.3	497
1972	13.5	8.9	4.6	12.0		687
Japan						Billions Yen
1954–56				13.0		5,969
1964	19.7	5.9	13.8	17.3	5.1	19,149
1968–70	22.9	9.8	12.8	19.7	5.5	40,796
1972	29.2	12.1	17.1	20.6	6.0	61,787

[a] Includes private households, personal trusts, and private nonprofit organizations.

[b] Aggregate disposable income was calculated from U.N. National Accounts statistics by subtracting the following from "Current Receipts of Households": all direct taxes including income tax, compulsory fees, fines, and penalties, social security contributions, and current transfers to either nationals or to the rest of the world.

Sources: Various issues of the following sources were used: Federal Reserve Bulletin; Bundesbank Reports; German Statistical Office National Account Series; Financial saving figures for other *OECD* Financial Statistics; U.N. Yearbook of National Accounts for 1973, German Statistical Office, National Account Series; Statistical Abstract of the U.S.; National Income and Expenditure: 1963–73 (by the Central Statistical Office of Great Britain); Japan Economic Yearbook.

statistical problems connected with cross-national comparisons of saving. Total saving (col. 4) is measured as a residual, namely as the difference between disposable income and total disbursements by private households, and is thus subject to inaccuracies and errors in measuring either of the two components. Total saving differs from the much more reliable financial saving statistics (cols. 1 to 3) in that it comprises undistributed profits or investments in own unincorporated business. Net investment in consumer durables and housing does not enter into either of these figures. While the other saving statistics in Table 1 represent the balance of saving over dissaving, Social Security contributions are gross figures, i.e., not adjusted for benefit payments and thus cannot simply be added to either col. 1, 3, or 4.

Table 1 leads to the following conclusions:

(1) Both total saving and net financial saving have been consistently much higher in Germany than in the United States. This is so in spite of considerably more "forced saving" through the German Social Security system resulting in considerably more income continuity extending beyond the working years.

(2) There is great similarity in gross financial saving between Western countries (excepting Japan). The important differences in net financial saving enter through the item increase in consumer debt which has been extremely high in the United States.

(3) Social Security contributions differ drastically, separating Germany and France on the upper end from the Anglo-Saxon countries on the lower one.

(4) Rates of saving in the United States show more stability over time, continuing a long history (R. W. Goldsmith). European and Japanese rates, while more volatile, have increased somewhat during the 1950's and 1960's.

In the following, the higher saving performance of Germany compared to the United States is viewed largely as resulting from:

(A) a greater felt need for financial reserves providing protection for contingencies and rainy days, brought about by charactistically pessimistic expectations for the future;

(B) a perceptible trend toward satiation with consumer goods making it easier for a saving motive of given strength to be translated into behavior;

(C) a policy of subsidizing mass saving, thus strengthening the instrumentality of saving for medium and long-term accumulation; and

(D) the ingrained practice of saving in advance of purchasing real estate and consumer durables rather than incurring installment debt, thus making saving highly instrumental for acquiring consumer investment goods.

(A) The link between cyclical changes in attitudes of mass publics (perceptions and expectations about own economic situation and that of the economy) and the volume of consumer durables purchases had first been demonstrated for the United States by G. Katona and later replicated for Canada and Germany (H. T. Shapiro and G. E. Angevine; Strumpel, K. Novy, and M. S. Schwartz). Optimism tends to induce consumers to spend more because they may be more willing to borrow against future income flows or to reduce their level of or addition to precautionary savings relative to income, tending to perceive past accumulations of reserves as adequate or redundant (W. E. Dunkelberg). A persistent posture of economic pessimism in Germany in the face of prosperity, sharply contrasting with the mood of confidence characteristic, until recently, for American households, was reported by Katona, Strumpel, and E. Zahn, who conducted a cross-national comparison of household attitudes and economic behavior on the basis of national surveys in 1968. Close to half (43 percent) of all American heads of household anticipated being better off four years from now, but only a third of the Britons, a quarter of the French, Germans, and Dutch. Whereas every third American both felt better off than four years ago and anticipated being better off four years hence, only every fourth Briton, every fifth Frenchman, every sixth Dutchman, and every eighth German felt similarly embedded in a continuum of progress. The Germans by far most frequently among Western nations denied that they were living in an extraordinarily favorable era of rising real incomes. Responding to unexpected income stagnation, Americans have only recently abandoned their traditional optimism and are now approaching German levels of consumer sentiment.

(B) Insecurity about the economic future breeds an apprehensive posture toward consumption and spending. Higher living standards are a byproduct of progress on which Germans tend to look with reserve, almost with suspicion. The maintenance and preservation of past gains are foremost in people's minds rather than any longing for better living. In 1957, 48 percent of German adults stated flatly that they would be satisfied if their economic situation would remain as it was for the next five to ten years. This proportion rose to 60 percent in 1961 and to 70 percent in 1963. When asked whether they considered the present average standard of living in Germany to be too high, too low, or just right, the proportion of those who disapproved of prevailing levels of consumption as being too high increased from 52 percent in 1958 to 62 percent in 1963.

Thus it is not surprising to find consumption aspirations of Germans lagging behind accomplishments. A symptom of the lag is an uninterrupted, steep increase in the proportion of households which are close to feeling "saturated." Asked whether respondents intended to make large outlays during the coming year, the negative answers increased from 16 percent in 1956 to 20 percent in 1959, 31 percent in 1962, and 46 percent in 1967. While comparative survey data for the United States are lacking, social observers emphasize the terse race between aspirations and accomplishment.

According to R. G. Merton, "In the American dream there is no final stopping point. . . . At each income level . . . Americans want just about twenty-five percent more (but of course this "just a bit more" continues to operate once it is obtained)."

(C) Since the currency reform of 1948, Germany has developed by far the most comprehensive system for promoting saving by both private households and enterprises. Within certain limits per household, tax deductions or outright subsidies were granted for time deposits, contractual saving through building and loan associations (Bausparkassen), acquisition of certain stocks and bonds, and live insurance premiums. There are also government incentives for allocating part of workers' wages or salaries to asset formation. Together with the normal interest payments, these premiums (some of them discontinued) added up to an effective annual interest rate of 10 to 15 percent for the amounts in question and thus by far exceeded market rates. The variety of promotion programs led to a high degree of habitualization and contractualization of saving behavior and thus tended to relieve the motivation to save from the continuous competition with instant consumption.

Moreover, the high effective rate of interest accruing on the accounts amounts to large quasi-automatic increments in liquid assets. A given original saving effort has resulted in more asset formation than in the United States or other countries lacking an extensive system for promoting saving.

(D) There are large cross-cultural differences in the purposes of saving. In Germany, a large proportion of liquid assets is intended for purchases of consumer durables. While 59 percent of the German savers interviewed in 1967 mentioned such purchases, only 7 percent of Americans in 1966 referred to the same items (Katona, Strumpel, and Zahn). In the United States in the late 1960's, close to one-half of all families had some installment debt, and close to 40 percent of all families bought something on the installment plan each year in the 1960's. The best figure available for Germany is that in the late 1960's 11 percent of all families were making monthly installment payments. This proportion did not increase at all in the 1960's and, according to some surveys, may even have declined somewhat.

In the United States, purchasing a home requires only a small down payment. Buying a home in Germany is contingent on a larger proportion of cash, much of it generated by long-term advance saving in building and loan associations.

The difference between saving in advance of a purchase and saving afterwards through repaying debt does not necessarily result in a difference in overall saving rates. If in the United States incurrence of new installment debt were balanced every year by repayment of such debt, and if in Germany the amounts saved in advance were balanced by withdrawals from liquid funds to make cash purchases, the overall saving rates of the two countries would not be affected by the different practices. The fact is, however, that in the last ten years in Germany families have accumulated liquid funds beyond the amounts they have used while in the United States outstanding installment debt has risen year after year far beyond the amounts of repayments made. Thus in a growing economy, the mode of financing consumer durables does influence the rate of saving. On the other hand, a decrease in the growth of mass incomes as expected by many would tend to reduce the differential.

The substantial differences in both goals and uses of saving are not just of a short-lived, cyclical nature. The peculiar manner in which Americans on the one hand and Germans on the other have interpreted their respective economic experiences and have used their savings and their credit has persisted at least through the 1950's and 1960's when rapid economic growth was accompanied by moderate inflation. How could a period characterized by slow growth or stagnation combined with high rates of inflation affect the motivation to save? On the one hand, disappointment or bad times might invoke the need to prepare for a dim and insecure future. Goals for saving would become stronger. On the other hand, rapid inflation would drastically reduce the actual and perceived instrumentality of financial reserves: accumulation-oriented financial saving presupposes trust in the stability of institutions and a certain predictability (not necessarily stability) of the value of money. The data exhibited in Table 2 suggest that during the highly inflationary year 1973 many Italians and Frenchmen for the time being have written off saving as an appropriate vehicle for transferring purchasing power from the present to the future. The respective saving figures for that year are not yet available. Germans, in the face of a more moderate increase in inflation, maintain their highly saving-minded posture and—as the

TABLE 2—ATTITUDES TOWARDS SAVING IN A PERIOD OF INFLATION

(In percent of interviewed households)[a]

	Germany		France		Italy	
	1973	1974	1973	1974	1973	1974
Considering the general economic situation:						
Saving makes (perhaps) sense	68	65	60	47	66	40
Saving makes (probably) no sense	29	29	32	47	27	49
Don't know	3	5	8	7	7	11

[a] The interviews were taken in January. The sample size exceeds 5,000 for each of the three countries. The sample is representative for all private households. Figures do not always add up to 100 percent because of rounding.

Source: Information P-30 of April, 1974 by the Commission of the European Communities, Brussels. The surveys are commissioned and coordinated by European Communities.

recent record shows—manage to translate it into behavior. Also Americans, consistent with the drastic deterioration in economic sentiment, have stepped up saving recently. As in the past twenty-five years, the security-deficit created by inflation superseded the decline in instrumentality, i.e., the fear of erosion of the value of savings (see the preceding paper by F. Thomas Juster and Lester D. Taylor).

In the quarter century of affluence following World War II, Western industrial nations went through a historically unique period of rapidly rising mass incomes. During this time span, monetary reserves ceased to be a prerogative of the rich. The real asset value of the average household, both financial and tangible, has increased substantially. This is true primarily for the United States, constant rates of financial saving notwithstanding. The advent of financial discretion or latitude made it possible for the many to extend their sights to the future, to prepare for expected and unexpected needs through reserve formation rather than to live from hand-to-mouth and from day-to-day. To provide for the future became a powerful aspiration of the masses albeit of lesser priority than minimum material subsistence.

The actual trends in these countries then appear to be in accordance with at least two basic tenets of the theories of the consumption function. The absolute income hypothesis tends to view saving as a function of ability to afford it; it thus manages to capture the hierarchy of priorities as suggested by the recent saving history. Yet it fails to consider saving and its dynamics as dependent on concrete motives whose strength may not vary mechanistically with income. The permanent income and life-cycle hypotheses, while negating any motivational priority of minimal present consumption over future needs, are explicit about the goal of saving. In centering on the desire for continuity of consumption levels, they describe one powerful motive of income allocation in industrial societies. Yet the individual or collective goal of equalizing consumption over time continues to compete with immediate consumption and other saving motives and remains a lofty vision for many. For instance, poverty even in affluent societies is still heavily concentrated among the old, in some countries more, in others less. And the latter are by no means the richer ones, as would have to be predicted on the basis of the absolute income hypothesis.

The failure of the theories of the consumption function to satisfactorily explain cross-national differences in the rate of saving has two reasons. First, with respect to the dependent variable, the theories neglect the multiplicity of saving goals. Saving is a conceptual and statistical artifact; its components differ with respect to their behavioral and motivational sources and merit separate theoretical attention. Second, with respect to the explanatory variables, the theories deal exclusively with income and its changes. This paper suggests that institutional and cultural characteristics of a society largely shape the uses and functions of saving and thereby its motivational makeup.

REFERENCES

W. C. Dunkelberg, "Forecasting Consumer Expenditures with Measures of Attitudes and Expectations," unpublished doctoral dissertation, Univ. of Michigan 1969, 125.

R. W. Goldsmith, *A Study of Saving in the United States,* Princeton 1955, *1.*

D. W. Johnson and J. S. Y. Chiu, "The Saving-Income Relation in Under-developed and Developed Countries," *Econ. J.,* June 1968, *78,* 321–33.

G. Katona, *The Mass Consumption Society,* New York 1964.

———, **B. Strumpel, and E. Zahn,** *Aspirations and Affluence: Comparative Studies in the United States and Western Europe,* New York 1971.

R. G. Merton, *Social Theory and Social Structure,* New York 1966.

E. Noelle-Neumann, "Geldwert und Oeffentliche Meinung," in C. A. Andreae et al., eds., *Geldtheorie und Geldpolitik, Essays in Honor of Guenter Schmoelders,* Berlin 1968.

H. T. Shapiro and G. E. Angevine, "Consumer Attitudes, Buying Intentions, and Expenditures: An Analysis of Canadian Data," *Can. J. of Econ.,* May 1969.

B. Strumpel, K. Novy, and M. S. Schwartz, "Consumer Attitudes and Outlays in Germany and North America," paper presented to the ninth Conference of *CIRET* (Int. Conf. of Business Tendency Surveys), Madrid, Sept. 1969, in: *Jahrbuch für Sozialwissenschaft, 1,* 1970.

Postwar U.S. Consumption, Consumer Expenditures, and Saving

By Michael R. Darby*

If interest in research topics ebbs and flows, there can be no doubt that macroeconomic consumption theory is again on a rising tide of fashion. This revival is the inevitable result of the greatly extended base of high quality data and of the central role of the consumption function in the debate between the dynamic and Keynesian approaches to macroeconomic theory.

The consumption theories developed in the 1950's by Milton Friedman and by Franco Modigliani and Richard Brumberg departed from that of Keynes in two significant ways: (1) the substitution of wealth for current income as the crucial determining variable, and (2) the substitution of consumption of service flows for consumer expenditures as the appropriate dependent variable. While this represented a substantial improvement over the Keynesian consumption function in the ability to explain the broad movements in consumption, it moved away from the key macroeconomic concept of consumer expenditures. Models must be developed which are appropriate to this latter concept.

Any successful explanation of consumer expenditures requires a clear distinction between consumption per se and consumer expenditures. This distinction is not "a theoretical nicety, inappropriate to empirical work."[1] Its neglect has led to some very large biases and bitter controversy over empirical magnitudes. I believe that the distinction is more than just empirically important. It is basic to a theoretical understanding of the transmission of monetary and other macroeconomic shocks.

I. The Relationship Between Consumer Expenditures and Consumption

Consumer expenditures, c_t^x, are quantitatively equal to consumption of service flows, c_t^f, plus the change in the stock of durable goods, Δd_t, less the implicit interest yield on durable goods, rd_{t-1}:[2]

$$c_t^x = c_t^f + \Delta d_t - rd_{t-1}. \qquad (1)$$

So there are three components to be explained by a theory of consumer expenditures: pure consumption, durables adjustment, and durables yield.

As is well-known, over long periods each of these components can be approximated as proportional to wealth or permanent income. In the short run, however, this approximation is unacceptable for the latter two components.

---

* Associate Professor, University of California, Los Angeles, and Harry Scherman Research Fellow, National Bureau of Economic Research, New York. Helpful suggestions were received from Armen Alchian, Phillip Cagan, and Arthur Gandolfi.

[1] M. J. Farrell (p. 679) complained of neglect of this distinction in 1959. It is present in the recent review article by Robert Ferber and the book by Thomas Mayer.

[2] All of these variables are measured in real terms. The real durable goods stock, d_t, is measured at the end of period t (beginning of period t+1). Equation (1) follows from the fact that $c_t^f = c_t^n + (r+\delta)d_{t-1}$ where c_t^n is the flow of purchases of nondurable goods and services, r the real interest rate appropriate to durable goods, and δ is the depreciation rate. Substitution in equation (1) yields $c_t^x = c_t^n + \delta d_{t-1} + \Delta d_{t-1}$ as required.

Pure consumption of service flows is properly assumed to be a fraction k of permanent income, y_{Pt}.[3]

$$c_t^f = ky_{Pt}. \tag{2}$$

Substituting equation (2) in equation (1) yields

$$c_t^x = ky_{Pt} + \Delta d_t - rd_{t-1}. \tag{3}$$

The stock of durable goods is the sum of past changes in the stock; so the explanation of short-run fluctuations in consumer expenditures concentrates on those changes.

II. The Estimation of Permanent Income

Before presenting an explicit model of durables stock adjustment, it is appropriate to comment briefly on implications of this approach for the construction of permanent income estimates. I (1974) have published a study of these implications elsewhere.

The permanent income hypothesis asserts that transitory income will alter pure consumption only in so far as it alters wealth. This implies that transitory income will be correlated with fluctuations in consumer assets including the stock of consumers' durable goods.[4] Neglect of this correlation between changes in the stock of durables and transitory income has led to specification error. The error arises in regressions which explain consumer expenditures—not pure consumption—by permanent income alone with transitory income omitted.

The standard analysis of bias due to omitted variables is inapplicable because permanent income is simultaneously estimated by an auxiliary equation

$$y_{Pt} = \beta y_t + (1 - \beta)(1 + g)y_{Pt-1}, \tag{4}$$

where β is the weight of current income, y_t, in the determination of permanent income and g is the trend rate of income growth. Selection of β so as to minimize the residual variance of the consumer expenditures regression, which is the standard procedure, leads to an upward bias of the β estimate.

Two interesting facts arose from my research on this statistical bias. (1) Substantial amounts of durable goods are officially classified as nondurable goods and services. These—such as clothing and shoes—biased upwards the weight of current income in permanent income even in studies such as Friedman's which replaced "durables" expenditures with their service flows. The estimated weight of current income in permanent income was reduced to about 0.1 on an annual basis. (2) The traditional permanent income estimator is derivable under the permanent income hypothesis from a perpetual inventory model of wealth. In that case the weight of current income in permanent income is the expected real yield on private wealth. The differences between permanent income and life-cycle models are thus primarily differences in techniques of estimating wealth.

III. A Consumer Expenditures Model

I will now complete the model of consumer expenditures by specifying a portfolio adjustment model of changes in the stock of consumer durables. By separating the pure consumption and portfolio adjustment decisions, many severe analytical hurdles are bypassed.

This research is still under way, but is sufficiently advanced that a progress report may be of interest. I am currently testing a generalized stock adjustment model of changes in the durables stock.

---

[3] There are several ways to derive this proportionality from maximizing behavior. Perhaps the simplest assumptions are a constant real yield r_p and infinitely-lived individuals with utility function homothetic in present and all future values of consumption.

[4] Empirical evidence for this proposition is presented in Darby 1972.

Changes in the durables stock are determined as the sum of four components: the normal change, a fraction of the difference in desired and adjusted actual stock of durables, a fraction of the difference between actual and permanent income (transitory income y_{Tt}), and a fraction of the difference between actual and desired real money balances.

$$(5)\quad \Delta d_t = (\Delta d_t)^e + \lambda_1(d_t^* - (\Delta d_t)^e - d_{t-1}) + \lambda_2 y_{Tt} + \lambda_3(m_t - m_t^*).$$

The first component of normal or planned change refers to the growth in the durable stock that would be associated with normal saving and growth in wealth. This amount would occur in the absence of any surprises and shocks.

The second component is a standard stock adjustment formulation adjusted for the existence of normal growth. This permits the long-run demand factors summarized by d_t^* to affect the rate of change of the stock.

The precise interpretation of the third component depends upon the view of permanent income and transitory income which is adopted. On the view taken in my restatement of the permanent income theory (1974), transitory income is unanticipated and therefore interpreted as a windfall change in wealth. Since it is received in nonhuman form, however, there will be a temporary change in all nonhuman consumer assets over and above that implied by the change in wealth alone. On the alternative view that transitory income is anticipated, it represents no change in wealth, but will be offset by future and past receipts of transitory income. So transitory income is saved in various asset forms. The first interpretation now seems to me the preferable one, and that approach is taken here.[5]

The final component—the excess demand for money—explicitly accepts the substitution explanation of the real balance effect. This specification is not restrictive empirically, however, since any wealth effect will increase the estimated coefficient of the real money supply without otherwise affecting the estimates.

In order to empirically estimate this equation, the normal change in the durables stock $(\Delta d_t)^e$, the stock demand for durables d_t^*, and the demand for money m_t^* must be specified. The stock demand for durables is assumed to be explained by a linear demand function in permanent income, relative price of durables P_{Dt}/P_{NDt}, and the long-term interest rate i_t:

$$(6)\quad d_t^* = \alpha_0 + \alpha_1 y_{Pt} + \alpha_2 \frac{P_{Dt}}{P_{NDt}} + \alpha_3 i_t.$$

The normal change in the durables stock is approximated as a constant times permanent income:[6]

$$(7)\quad (\Delta d_t)^e = \eta y_{Pt}.$$

Finally, the demand for money is taken as a function of permanent income, transitory income, and the long-term interest rate:

$$(8)\quad m_t^* = \gamma_0 + \gamma_1 y_{Pt} + \gamma_2 y_{Tt} + \gamma_3 i_t.$$

Combining equations (3) and (5) through (8) yields the complete model of consumer expenditures:

---

[5] This interpretation implies that transitory holdings of durable goods will be eliminated over time even in the absence of offsetting transitory income receipts. The alternative view was taken in Darby (1972). The view that transitory income increases or decreases wealth dollar for dollar is subject to criticism within the context of a neoclassical growth model. Many would object to the implied endogenicity of labor due to investments in human capital.

[6] Note that $\Delta d_t^* = \alpha_1 \Delta y_{Pt} + \alpha_2 \Delta(P_{Dt}/P_{Nt}) + \alpha_3 \Delta i_t$. If there is no expected trend in the two-price terms, $(\Delta d_t^*)^e = \alpha_1(\Delta y_{Pt-1})$, the expected growth in durables demand can be so approximated. Current permanent income is substituted for lagged permanent income in order to avoid problems of multicolinearity.

$$(9)\quad c_t^{x} = \beta_o + \beta_1 y_{Pt} + \beta_2 y_{Tt} + \beta_3 m_t + \beta_4 d_{t-1} + \beta_5 \frac{P_{Dt}}{P_{NDt}} + \beta_6 i_t + u_t.$$

On the derivation presented above, definite signs can be assigned only to β_3 (positive) and to β_4 and β_5 (negative), though there is a presumption that the negative impact of permanent and transitory income in increasing the real demand for money will not fully offset the otherwise positive impact.

An advantage of estimating equation (9) is that the form—though not the interpretation of coefficients—is robust with respect to several different ways of modeling the durable stock adjustment equation. Also the fact that the official definition of durable goods excludes substantial amounts of behaviorally durable goods is finessed to the extent that the stock of these quasi-durable goods moves similarly to the stock of the officially designated durable goods.

IV. Estimation of the Consumer Expenditure Function

The consumer expenditure function (9) may shed light on several outstanding empirical questions: (1) Does the estimate of 0.1 as the weight of current income in determining permanent income which was estimated on the basis of consumption flows hold up for consumer expenditures? (2) What is the proper definition of current income for use in explaining consumer behavior? (3) Which definition of the money supply seems most appropriate for the consumer expenditure function?

The first question has considerable personal interest, since it deals with the general applicability and consistency of the approach contained in my 1974 restatement. That question must necessarily be dealt with in the simultaneous estimation of equations (4) and (9).

The second question has been frequently raised of late. It is illogical that consumers should pay much attention to disposable personal income regardless of variations in saving on their behalf on the part of firms. Many researchers have found that the addition of this saving (undistributed corporate profits) to disposable personal income substantially improves their ability to explain consumer behavior. From the point of view of completeness, a more thoroughgoing definition of private sector income would be net national product less taxes net of transfers. This series should be adjusted for the yield on the stock of consumers' durable goods. From the point of view of a rational consumer, the inflationary tax on real base money should also be deducted in deriving the amount which can be allocated between consumption and increasing wealth.[7]

The third question refers to the possibility of using consumer expenditure data to provide information on the debate in monetary theory about the empirical definition of money. Similarly, estimation of permanent income on the basis of the demand-for-money equation may provide some further information on the weight of current income in the determination of permanent income.

These three issues must finally be considered simultaneously. Unfortunately the empirical investigation is still quite incomplete. Suggestive estimates are available, however, for a single combination of real income and real money stock definitions. Real private sector income adjusted for a 10 percent yield on the stock of consumer's durable goods and the narrow (M_1) money supply are the concepts used. The real stock of consumers' durable goods is extended from Darby 1972. The relative price of consumers' durable goods is based on the implicit price deflators indicated. The interest rate i_t is the market yield on long-term *U.S.* government bonds. All variables except the last two are measured in billions of 1958 dollars.

---

[7] Arguably the rate of inflation times at least part of the real government debt should also be deducted.

For annual data from 1947 through 1973, the estimated consumer expenditure function is:[8]

$$
\begin{aligned}
(10)\quad c_t^x = & -\underset{(57.90)}{143.6} + \underset{(.065)}{1.07}y_{Pt} + \underset{(.060)}{0.41}y_{Tt} \\
& + \underset{(.135)}{0.66}m_t - \underset{(.069)}{0.36}d_{t-1} \\
& + \underset{(36.3)}{27.3}\,\frac{P_{Dt}}{P_{NDt}} + \underset{(1.35)}{1.44}i_t \\
& \text{S.E.E.} = 1.996 \quad R^2(\text{adj.}) = .9996 \\
& D - W = 2.37
\end{aligned}
$$

The maximum likelihood value of β is 0.15 (to the nearest 0.01) with a greater than 90 percent probability confidence interval (0, 0.23).[9] The corresponding quarterly regression is:

$$
\begin{aligned}
(11)\quad c_t^x = & -\underset{(8.80)}{27.2} + \underset{(.032)}{0.87}y_{Pt} + \underset{(.036)}{0.45}y_{Tt} \\
& + \underset{(.025)}{0.19}m_t - \underset{(.009)}{0.034}d_{t-1} \\
& + \underset{(5.59)}{2.87}\,\frac{P_{Dt}}{P_{NDt}} + \underset{(0.22)}{0.38}i_t \\
& \text{S.E.E.} = 0.744 \quad R^2(\text{adj.}) = .9992 \\
& D - W = 1.07
\end{aligned}
$$

The maximum likelihood value of β is 0.005 (to the nearest 0.0025) with a greater than 90 percent probability confidence interval (0.0, 0.06). In comparing these two regressions, note that the flow variables c_t^x, y_t, y_{Pt}, and y_{Tt} are measured for the quarterly regression at seasonally adjusted *quarterly rates* in view of the stock-flow relationships.

A few comments on these results are in order. First the explanatory power: the ratio of the standard error of estimate to the mean is 0.59 percent for the annual data and 0.86 percent for the quarterly data. The two regressions correspond closely when the coefficients are adjusted for the different time units. The two stock variables—real money and the durable goods stock—enter statistically significantly and with the expected signs. The estimated values of the weight of current income in permanent income bracket the value of 0.10 per annum estimated in the previous work, though the differences are not statistically significant.[10] The coefficient of the relative price of durable goods P_{Dt}/P_{NDt} is statistically insignificant and of the wrong sign. This is not a surprising result since the prices of about half of all durable goods—in a behavioral sense—are included in the denominator because of misclassification. The significantly negative constant term is due to the positive constant in the demand for money.

The consumer expenditure function also provides a useful avenue to analyze saving behavior. The concept of saving which is directly applicable is private sector saving (private sector income less consumer expenditures). Again, this is the concept of prime interest to macroeconomists. It is trivial to show that private saving is determined by:

$$
\begin{aligned}
(12)\quad s_t = & -\beta_o + (1-\beta_1)y_{Pt} + (1-\beta_2)y_{Tt} \\
& - \beta_3 m_t - (\rho + \beta_4)d_{t-1} - \beta_5\frac{P_{Dt}}{P_{NDt}} \\
& - \beta_6 i_t - u_t,
\end{aligned}
$$

---

[8] Standard errors are given in parentheses below the coefficients.

[9] Computed on the basis of the asymptotic distribution of the *log* of the likelihood function. The likelihood function is rather flat between 0 and 0.2, after which it drops rapidly.

[10] In further work, it is found that this weight is closer to 0.1 if the years 1947 through 1953 are dropped from the regression. Whether this is due to peculiarities in the data—such as those due to price controls—or in the initial values selected for the permanent income or durables stock is unknown.

where the β_i's are as in equation (9) and $\hat{r}$ is the imputed yield on the durables stock.[11] Regression fits of equation (12) correspond with regressions (10) and (11). The model explains private saving very well, with a standard error of estimate only 5.0 percent of the mean for annual data and 7.5 percent for quarterly data.

The long-run stability of the ratio of private saving to private sector income does not hold in the short run because of fluctuations in the levels of transitory income, real money balances, and the real durables stock. There is no reason to expect stability of the ratio of personal saving to personal disposable income over any period.

V. Summing Up

If the research does not yet permit conclusions, it does suggest a point of view. The long-run pattern of consumer expenditures conforms to a trend-dominated measure of permanent or normal income. Short-run deviations of consumer expenditures reflect three main influences: (1) transitory income, (2) the real money supply, and (3) the real durables stock. The first measures the current rate of change in nonhuman assets relative to normal saving. Given permanent income and transitory income, real money balances enter as a measure of the excess money supply. Higher real money balances have a very clear and economically significant effect on directly increasing real consumer expenditures while lower real money balances directly decrease real consumer expenditures.[12] The level of the real durables stock has a definite negative impact on the rate of consumer expenditures due both to their implicit yield and to stock adjustment.

There is much work yet to be done to examine the impact of the alternative definitions of income and money on the estimated weight of current income in permanent income. Statistical problems of simultaneity, auto-regressive residuals, and suspect 1947–53 data also arise.

It seems to me that a strong case has been made, however, for formulating aggregate consumer expenditure functions instead of consumption functions. The general approach of equations (1) and (4) provides a broad framework for other researchers to develop improved models of the adjustment of the durables stock or even of pure consumption. This combined approach is less sensitive than attempts to convert the data into pure consumption with respect to what are little more than arbitrary decisions as to what is and what is not a durable good. And, of course, consumer expenditure functions have the advantage of explaining the variable of interest to macroeconomists.

REFERENCES

M. R. Darby, "The Allocation of Transitory Income Among Consumers' Assets," *Amer. Econ. Rev.*, Dec. 1972, *62*, 928–41.

———, "The Permanent Income Theory of Consumption—A Restatement," *Quart. J. Econ.*, May 1974, *88*, 228–50.

M. J. Farrell, "The New Theories of the Consumption Function," *Econ. J.*, Dec. 1959, *69*, 678–96.

R. Ferber, "Consumer Economics, A Survey," *J. Econ. Lit.*, Dec. 1973, *11*, 1303–42.

M. Friedman, *A Theory of the Consumption Function*, Princeton 1957.

T. Mayer, *Permanent Income, Wealth, and Consumption*, Berkeley 1972.

F. Modigliani and **R. Brumberg**, "Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data," in K. E. Kurihara, ed., *Post Keynesian Economics*, New Brunswick 1954

---

[11] That is, 0.1 per annum for annual data and 0.025 per quarter for quarterly data.

[12] This provides strong support for the monetarist proposition that changes in the money supply shift the *IS*—as well as the *LM*—curve.

DISCUSSION

Thomas Mayer, University of California, Davis: Michael Darby's paper is a very competent piece of work. He uses standard econometric techniques in a sophisticated way, but still I am uneasy about his results. This is so primarily because I feel uneasy about the validity of our standard techniques, not because of the particular way he uses them.

One difficulty with standard techniques appears when one asks what is new about Darby's results: In what way do they differ from the standard view? His use of both permanent income and transitory income is fairly standard by now. Moreover, his finding that the coefficient of permanent income is about twice the coefficient of transitory income is consistent with the bulk of the empirical evidence. Specifically, it is consistent not only with a modified permanent income theory such as Darby presents, but also with a theory intermediate between the permanent income theory and the older theories, a theory which has households consuming both out of permanent and transitory income directly. Similarly, his inclusion of the money stock and the interest rate in the consumption function is also consistent with previous work. To be sure, the relative price variable is normally not used, but since it has the wrong sign, it really does not belong in Darby's function either. Finally, his *d* variable has been used by Hendrik S. Houthakker and Lester D. Taylor.

This raises the question whether the detour via the consumption of durable services is really needed. Darby could have written out the function he actually tested directly, based on a general theory of household behavior. It can be derived from a much broader range of theories than the permanent income theory. For example, in Darby's model the transitory income term enters as a part of the permanent income theory because some so-called nondurables are actually durables. But one could also say that households consume directly out of transitory income. Hence, one can accept Darby's equation without accepting his permanent income theory. This, of course, is a common problem in econometric work. When reduced to testable propositions, quite different hypotheses come up with the same equation and therefore cannot be tested against each other. However, one result of Darby's paper is new. This is the very low value of beta, and hence the very long lag in the formation of permanent income. But as discussed below the evidence for this is not very strong.

A second problem with time series regression tests is that many quite arbitrary decisions inevitably have to be made and poor proxies have to be used. For example, Darby simplifies by treating the desired stock of durables as proportional to permanent income. But why should the income elasticity of durables be equal to unity? Second, Darby explains the inclusion of the money stock as representing an allowance for a portfolio adjustment process. But why confine a portfolio adjustment process to money; why not include savings and loan shares, etc.? Third, the inclusion of the money stock creates a problem of reverse causation since an increase in consumption, by raising interest rates, is likely to cause the Federal Reserve and the banks to raise the money stock. (Obviously, the transitory income variable and the interest rate also have a potential for "reverse causation.") Fourth, the interest rate variable is measured badly. As Milton Friedman has rightly insisted, the observed interest rate on securities is a very poor measure of the expected real rate of return which households can obtain from holding a wide variety of nonmoney assets. (Darby uses a bond rate as his measure of the interest rate. A composite deposit rate on bank time deposits, savings and loan shares, mutual savings bank deposits, and consumer loan rates would have been better since, until recently at least, households did not buy bonds to any significant extent.) Fifth, since quality improvements have probably been more important for durables than for nondurables, the relative price variable is probably measured quite badly, and it is therefore not surprising that it has the wrong sign. (But upon finding that it has the wrong sign, Darby should have eliminated this variable and rerun his regression without it.) Sixth, Darby forms his permanent income estimate by using a maximum likelihood procedure in which the computer is allowed to select the beta coefficients which maximize R^2. But this commonly used—and perhaps only available—technique is of questionable validity. Recently I tried to test the validity of goodness of fit tests by comparing the ranking of various hypotheses by goodness of fit during the sample period with their ranking when projected beyond the sample period. And the results were not encouraging, to put it mildly. Finally, while I like Darby's adjustment for corporate savings, I am not sure that it goes far enough; perhaps other capital gains should be included too. The fluctuations observed in the price-earnings ratio suggest that corporate savings is not a reliable guide to expected capital gains.

Hence, there exist a large number of points on which Darby's paper can be criticized. But this does not mean that he did his econometrics badly. Unlike many discussants, I do not want to give the impression that had I been the author I would have avoided such mistakes and have done the job much better. This is *not* the case. The problems I have outlined result primarily from the absence of good proxies, and from other problems which are inevitable in econometric work. But since these problems do exist, I believe that Darby's results, while interesting and suggestive, are far from conclusive. The way to obtain more conclusive results is to run many tests using different sets of data and different techniques. If such tests all essentially tell the same story, then and only then can one conclude that a strong case has been made. And I welcome Darby's study as one step along this path.

Turning briefly to the F. Thomas Juster-Lester D. Taylor paper, here too existing techniques have been used very competently. But again, our standard regression techniques reveal a weakness. Adding the uncertainty variable raises the explained variance only slightly, though significantly so. (The standard error of estimate falls substantially. But since this error is very low in any case (only 6 percent of the mean of the dependent variable) it probably is not a good measure of the performance of the equation beyond the sample period.) To be sure, since the fit is already good prior to the addition of the uncertainty variable, one could hardly expect more, but even so, I am left with *some* doubt. But the uncertainty variable is one which is highly appealing on a priori grounds, and Juster and Taylor's paper is therefore very useful in showing that a variable which a priori seems to belong in the consumption function actually does improve the fit of the consumption function.

MICROECONOMIC THEORY: CONFLICT AND CONTRACT

A Contractarian Paradigm for Applying Economic Theory

By JAMES M. BUCHANAN*

The object for economists' research is "the economy," which is, by definition, a *social organization*, an interaction among separate choosing entities. I return deliberately to this element in our primer, because I think that it has been too often overlooked. By direct implication, the ultimate object of our study is not itself a choosing, maximizing entity. "The economy" does not maximize, and we may substitute "the polity" here without change in my emphasis. No one could quarrel with these simplistic statements. The inference must be, however, that there exists no one person, no single chooser, who maximizes *for* the economy, *for* the polity. To impose a maximizing construction on the models that are designed to be helpful in policy is to insure sterility in results.

Where did economics, as a discipline, take the wrong turn? My own suggestion is that Lionel Robbins marks a turning point. His book defined "the economic problem" as the location of maxima and minima. Almost simultaneously with this, the Edward H. Chamberlin and Joan Robinson books marked a turning inward, so to speak, a shift toward the maximizing problem of a specific decision-making entity. The economics of the firm was born, to be followed by the Hicksian elaboration of the economics of consumer choice. Paul Samuelson put this all together in his *Foundations of Economic Analysis*. Importantly, he extended the maximizing construction to welfare economics, extolling the virtues of A. Bergson's social welfare function as the tool through which such extension was made possible. For a quarter of a century, we have witnessed many variations on this theme, with economists hither and yon maximizing objective functions subject to specific constraints.

I should not imply that the maximizing models have held monolithic dominance. The institutional economists, and their successors, have continued their sometimes inarticulate critique of economic theory. Frank Knight, and some of his students, continued to lay stress on the social-organization aspects of the discipline. Game theory, in its solution rather than its strategy search, offered partial redirection of emphasis. More importantly for my purposes, public choice theory emerged as the positive theory of politics, a theory that necessarily treats individual decision takers as participants in a complex interaction that generates political outcomes.

But let me return to mainstream efforts of economists in the years since World

---

* University Professor and General Director, Center for Study of Public Choice, Virginia Polytechnic Institute and State University. The author is indebted to Amoz Kats and Gordon Tullock for helpful comments.

War II. I have no quarrel with the elaborations and refinements of the maximizing models for individual and firm behavior, although I have argued that many of these contributions belong, appropriately, to home economics or to business administration rather than to political economy. My strictures are directed exclusively at the extension of this basic maximizing paradigm to social organization where it does not belong. This is the bridge which economists should never have crossed, and which has created major intellectual confusion. "That which emerges" from the trading or exchange process, conceived in its narrowest or its broadest terms, is not the solution to a maximizing problem, despite the presence of scarce resources and the conflict among ends. "That which emerges" is "that which emerges" and that is that.

Return to game theory for analogy. The solution to a game with defined rules is not a maximum, and an external observer of the game would not attempt to seek improvements by operating directly on solutions. He would, instead, look at the prospects for changing the rules, and observed solutions would be information inputs in his evaluation of the game. Solutions are not directly "chosen" by a welfare function that embodies the preferences of the players, and solutions are not themselves ordered in terms of such a function. In game theory, the attempt to force such a construction would appear (and be) absurd. Yet this seems to me to be precisely what economists try to do within their analogous domain. Game theorists would indeed be surprised if someone should find that solutions could, in fact, be ordered in any manner that was consistent with the values of individual players. And they would not be profoundly shocked, to say the least, if someone should prove that such an ordering could not be made. Yet is this not what most economists experienced when Kenneth Arrow published his famous impossibility theorem in 1951?

My own initial reaction to Arrow's work was, and remains, one of nonsurprise (see Buchanan 1954). Who would have expected any social process to yield a consistent ordering of results? Only economists who had made the critical methodological error of crossing the bridge from individual to social maximization without having recognized what they were doing would have experienced intellectual-ideological disappointment.

There is no need to limit discussion to Arrow's theorem, although I shall return to this. Economists crossed the bridge from individual to social maximization because they wanted to be able to say something about policy alternatives. They desperately needed some instrument which would allow them to play the social engineer even if they eschewed the explicit intrusion of their own values in the process.[1] With the social welfare function construction, they could then talk as if their policy statements were operationally meaningful, and this ability provided them with a certain inner satisfaction. They have remained unwilling to utilize the Pareto criterion as a mere classification scheme, which Ragnar Frisch had advised them to do, and they have not followed my own suggestion (1962) about shifting the application of the criterion back to the level of institutional choice, where prospects for mutuality of agreement are enhanced.

How would the abandonment of the social maximization paradigm have changed the research thrust of economic theory, especially its potential relevance for social policy? The Pareto criterion classifies all positions into two mutually exclusive sets. Once an initial position is explicitly identified to be "nonoptimal" or "inefficient," we know that there exists at least one means of moving from this position to a position that falls within the optimal or efficient set. The initial position must be dominated, for all persons, by at least one position in the efficient set.

---

[1] I am indebted to Charles Plott for suggesting this explanation of the Bergson-Samuelson approach, which is a more sympathetic explanation than the one which I have elsewhere suggested.

There is no issue here, but what should be the role for the economist who has completed these first steps in applying his expertise? He should neither revert to nihilism nor seek the escapism of social welfare functions. His productivity lies in his ability to search out and to invent social rearrangements which will embody Pareto-superior moves.[2] If an observed position is inefficient, there must be ways of securing agreement on change, agreement which signals mutuality of expected benefits. In the limiting case, compensation schemes can be worked out which will achieve Wicksellian unanimity. Yet how many economists do we observe working out such schemes? How many economists bother with proposed compensations (which must, of course, include structural-institutional rearrangements) to those who will be overtly harmed by the effects of a public or governmental policy shift?[3] Instead of this potentially constructive effort, we find our colleagues continuing to express opposition to tariffs, quotas, minimum wages, price controls, depletion allowances, monopolies, tax loopholes, etc., whether these be existing or proposed. And they continue to stand surprised when the political process, as it operates, pays little or no heed to their advice. As Knut Wicksell noted eighty years ago, economists act as if they are advising a benevolent despot, in which case their post-Robbins, post-Bergson stance would, of course, be entirely fitting.

There is a place for efficiency in my suggested scheme of things. In this sense, there is little wrong with economic theory per se. Efficiency, as an attribute, is necessarily present when there is a demonstrated absence of possible agreed-on changes. The trading process, broadly conceived, is the means through which the "potentially realizable surplus" is exploited. But there is no uniquely determinate outcome of the trading process, since exchanges must be made and contracts enforced at preequilibrium and, hence, disequilibrium prices. Under the standard assumptions, simple exchange insures that an efficient or optimal position is attained, but this is only one from among a set containing a subinfinity of possible positions. Economists should be satisfied with this result. Within theory itself, search for uniqueness seems to be misguided effort. When the standard assumptions are not descriptive, the simple exchange process will not generate efficient results, and more complex arrangements may be called for. But these arrangements may still be examined in a contractual framework. The specification of these complex contractual arrangements may challenge the skill of the practicing political economist. In facilitating these complex exchanges, which may require the inclusive membership of the whole group, collective-governmental institutions may be necessary. And even though unanimous agreement for change may be conceptually possible, the costs of reaching agreement may be acknowledged to be prohibitively high. This suggests, in turn, that rules or institutions for reaching collective or group decisions may be preselected at some constitutional stage of "trade." It becomes possible in this way to apply the basic contractarian paradigm to the discussion of possible agreement on rules, even if it is anticipated that conflict will emerge at some final stage of application.

---

[2] This is essentially the suggestion that I made in my paper in 1959.

[3] An exception is W. H. Hutt. In a much neglected small book published in 1943, he proposed that the postwar British economy be swept clear of all market restrictions through the device of compensating all persons and groups who would lose by the change.

Whether the contractarian paradigm is applied at the level of simple exchange, within the constraints of well-defined rules, or at the most basic constitutional level where institutions themselves are the objects upon which agreement must be reached, or at any intermediate level, the emergent results of the trading process are properly summarized as a set of optimal positions, each one of which represents a possible outcome, and no one of which dominates any other in the set. This statement is enough to suggest my prejudices toward game-theoretic explorations into mathematics as opposed to the intricacies of complex maximization. The continuing search for solution concepts—the J. von Neumann-O. Morgenstern solution set of imputations, the several cores along with other more sophisticated concepts—seems to me to reflect and in turn to foster an attitude in the theorist that is consistent with, and contributory to, the contractarian approach that I am here suggesting.

The modern efforts to prove that competitive equilibrium exists and lies in the core of an economy are within my own limits for methodological legitimacy. The complementary emphasis here should be placed on the multiplicity of possible equilibria, on the absence of uniqueness. The devices of the Walrasian auctioneer or Edgeworthian perfect recontracting have been required for the proofs here, but I urge those skilled in this particular mathematics to search for theorems that require less stringent assumptions, that incorporate trading at disequilibrium prices.

Now let me return, as promised, to the Arrow impossibility theorem. In his Nobel Prize lecture, Arrow (1974) offers a lucid summary of modern economic theory, and he concludes with a discussion of his impossibility theorem. Unfortunately, Arrow does not seem to have gone beyond his initial failure to appreciate the inconsistency between his norms for a social welfare function, or choosing process, and the precepts for a society of free men. In his lecture, Arrow makes it clear that the Bergsonian requirement of "collective rationality" is prior to the set of reasonable conditions that he lays down for his social ordering. Having long ago proved that these conditions cannot be met, Arrow continues to hold that "the philosophical and distributive implications of the paradox of social choice are still not clear." This statement is both surprising and personally disappointing, since it indicates that Arrow has paid no heed to the arguments which I have made, along with many others, against the whole notion of collective rationality.

The so-called paradox may be used as a single and simple illustration of the profound difference between the maximization paradigm and the contractarian one. There are three voters, A, B, and C, and three alternatives, 1, 2, and 3. The collective choice rule is simple majority voting. Through a series of pairwise comparisons, individual preference orderings may be such that 1 is majority-preferred to 2, and 2 is majority-preferred to 3, while 3 is majority-preferred to 1. This cyclical result is disturbing only to those who seek uniqueness in outcome, who seek to impose the maximization paradigm on a social interaction process where it does not belong. By contrast, to those who accept the contractarian paradigm, who seek only to explain and to understand the behavior of persons who interact, one with another, there is nothing at all disturbing in the paradox. On the assumptions that all side payments have been made, or that institutional constraints are invariant, the tools of economics enable us to classify all three positions, 1, 2, and 3, as falling within the Pareto optimal or efficient set, and no one of these positions dominates the other. If individual preferences produce this set of results, we should be content with this and forego the essentially misleading searches for "philosophical implications" which simply are not there.

In this paper I have restated a position that I have presented in bits and pieces over almost two decades. Knight's favorite quotation from Herbert Spencer says that, "only by varied reiteration can alien conceptions be forced on reluctant minds." I hope that I have at least varied the reiteration sufficiently to avoid boredom. I would not presume to think that my own views on methodology would convince more than a minority of my professional colleagues, and I am under no illusions as to the continued dominance of the maximization paradigm in modern economics. Nonetheless, I think that progress has been made during the last twenty years. Game theory, after a series of disappointing attempts to work out optimal strategies for players, has shifted toward a more comprehensive, and more appropriate, consideration of solution concepts. Public choice has emerged as a subdiscipline in its own right, and one that is currently thriving, both within economics and in political science. But intellectual developments have perhaps been overshadowed in effect by the march of events. As persons, both from the streets and the ivory towers, observe modern governmental failures, they can scarcely fail to be turned off by those constructions which require beneficent wisdom on the part of political man. And they can hardly place much credence in the economist consultant whose policy guidelines apply only within institutions that embody such wisdom. Something is amiss, and economists are necessarily being forced to take stock of the social productivity of their efforts. When, as, and if they do so, they will, I think, come increasingly to share what I have called the contractarian paradigm.

As I have argued elsewhere (1964, 1969), economics comes closer to being a "science of contract" than a "science of choice." And with this, the "scientist," as political economist, must assume a different role. The maximizer must be replaced by the arbitrator, the outsider who tries to work out compromises among conflicting claims. The Edgeworth-Bowley box becomes the first diagram in our elementary textbooks; the indifference curve-budget line construction is relegated to subsidiary treatment. Eugen v. Böhm-Bawerk's horse traders are the basic examples, not the housewife who shops for groceries in the supermarket. Game theory, in its most comprehensive sense, becomes the basic mathematics for the professional, and solutions to n-person games replace the nth order conditions for maxima and minima. The unifying principle becomes *gains-from-trade*, not maximization. These principles merge, of course, at the level of the individual chooser's calculus, but they become quite distinct when attention shifts to the social interaction that we call "the economy."

REFERENCES

K. Arrow, *Social Choice and Individual Values*, New York 1951.

———, "General Economic Equilibrium: Purpose, Analytic Techniques, Collective Choice," *Amer. Econ. Rev.*, June 1974, *64*, 253–72.

A. Bergson, "A Reformulation of Certain Aspects of Welfare Economics," *Quart. J. of Econ.*, 1938, *52*, 310–34.

J. M. Buchanan, "Social Choice, Democracy, and Free Markets," *J. of Pol. Econ.*, 1954, *62*, 114–23.

———, "Positive Economics, Welfare Economics, and Political Economy," *J. of Law and Econ.*, 1959, *2*, 124–38.

———, "The Relevance of Pareto Optimality," *J. of Conflict Resolution*, 1962, *6*, 341–54.

———, "What Should Economists Do?," *Southern Econ. J.*, Jan. 1964, *30*, 213–22.

———, "Is Economics the Science of Choice?," in E. Streissler, ed., *Roads to Freedom: Essays in Honor of F. A. Hayek*, London 1964, 47–64.

E. H. Chamberlin, *The Theory of Monopolistic Competition*, Cambridge, Mass. 1932.

R. Frisch, "On Welfare Theory and Pareto Regions," *Int. Econ. Papers*, 1959, *9*, 39–92.

J. R. Hicks, *Value and Capital*, London 1939.

W. H. Hutt, *A Plan for Reconstruction*, London 1943.

L. Robbins, *An Essay on the Nature and Significance of Economic Science*, London 1932.

J. Robinson, *Economics of Imperfect Competition*, London 1933.

P. A. Samuelson, *Foundations of Economic Analysis*, Cambridge, Mass. 1947.

J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton 1944.

K. Wicksell, *Finanztheoretische Untersuchungen*, Jena 1896.

Problems vs. Conflicts: Economic Theory and Ideology

By DUNCAN K. FOLEY*

This paper is an attempt to think about the relation between individualist economic theory and social reality. This relation arises in two ways: first, scientifically, in the problem of explaining real phenomena; and second, in policy, in the attempt to intervene to alter social reality in the light of economic theory. I believe that individualist theory has some grave problems in both these areas, and I will try to make some suggestions as to why it continues as a theory in the face of these problems.

To begin with, let me try to mark off what I mean by individualist economic theory, which I use in preference to "neoclassical," because neoclassical has so many different meanings for different people. I use individualist to describe the notion that social reality as we experience it is in its important economic aspects a reflection of certain tastes or preferences of individual people interacting through processes of choice with objectively given technological possibilities, whatever particular market or informational structures are thought of as mediating the individual choices.

I should emphasize that I am not here trying to criticize all the work at all levels done by people who believe in individualist economics, much of which does not, in fact, use the individualist theory in explanation. My exercise is directed at the "high" theory, perhaps best expressed today in mathematical general equilibrium theory and welfare economics, but which also underlies the work of the Chicago school of libertarian economics.

I.

Let me first take up what I feel is the fundamental scientific problem with individualist economic theory. The aim of scientific theory is to provide objectively verifiable explanations of phenomena. These explanations often involve a reference to invisible structures, the existence of which is posited by theory and the properties of which are deduced from observations given the theory. The theory of physics, for example, explains many of the properties of matter as derived from properties of elementary particles like electrons. Although electrons are invisible, we believe that they exist because the theory of physics explains so many phenomena, and because properties of electrons, such as mass and charge, can be inferred from many different experiments all of which agree. If the mass and charge of electrons were constantly changing, the electron would not itself be a very powerful explanatory structure.

If we are to take individualist economics seriously as a scientific (that is, explanatory) theory, then I think we must interpret the concept of "tastes" in the same way as we understand the concept of the electron. "Tastes" or "preferences" have to be taken as structures that have a material reality but are not directly observable. (I leave "technology" out of this discussion entirely.) Individualist economic theory proposes to explain phenomena by reference to tastes: "*X* happened in part because tastes were of a certain kind at the time." In this situation the explanatory power of the theory depends on identifying stable properties or laws of evolution of tastes. If tastes have no stable properties, then we learn nothing from an explanation that refers to them, and there is not much scientific reason even to believe in their material existence.

---

* Stanford University. I would like to thank Jens Christiansen, Partha Dasgupta, Paul David, Bridget O'Laughlin, Mike Spence, Dave Starrett, Joe Stiglitz and the Political Economy Seminar at Stanford University for discussions that helped me greatly in writing this paper. The ideas expressed are my own and I suspect that every person in the list above would disagree with at least one of them. In addition, I owe a considerable debt to many other people who also helped me on the road to the method of analysis and point of view that lie behind this paper. Among them are Mike Carter, Jack Gurley, Don Harris, Rod Jacobs, Chiranjib Sen, Gita Sen, and Sandy Thompson.

One would expect, then, that a central topic of individualist economics would be the identification of stable properties of tastes and of the laws that govern their evolution. It is striking that so little effort goes into this problem. It is not enough from this point of view to specify tastes that are consistent with a given set of demand data; some properties must be proposed that are stable over time and space or that evolve according to determinate laws. Measurements from different places and times would have to support each other directly or support a general law of tastes before we can begin to believe in tastes as an invisible but material determinant of social reality or find explanations based on tastes convincing.

I have come to doubt whether this program of establishing the material reality and properties of tastes can ever be carried out. Words like "prefer" and "choose," which are basic concepts of individualist theory, basically are used to express how a person subjectively experiences a situation. When one describes a situation as one of "choice" or "coercion," one is saying something about oneself rather than about the situation, since the same objective behavior can be experienced either as choice or coercion (or as "preference" or "habit" and so on). It seems to me questionable to include such subjective categories in objective explanations or descriptions of social reality. I think it is even an error to think of the problem as one of what determines tastes, that is, to seek objective factors that influence preferences, which then can be viewed as determining behavior. In this kind of scheme, the tastes themselves become superfluous. It seems rather that the question of how people experience a situation as subjects is a question on a totally different level from the question of what their behavior is.

Individualist economic theory appears to remain stuck in a particularly old-fashioned and unsatisfactory view that all human beings experience themselves as rational, calculating entities weighing certain ends against available means and experience their behavior as rational choice. Aside from the fact that this appears to be a postulate empty of explanatory power, it blinds individualist theory to the differences between people as they enter into relations with each other, for example, in the labor market, to the objective differences between the worker, who has nothing but his labor power and enters into the exchange in order to survive, and the capitalist, who controls his own and the worker's means of subsistence and enters into the exchange to make a profit. Both may (or may not) experience themselves as being rational decision makers, but that fact doesn't get us much further in understanding the nature of the relation they enter into (is it, for example, an equal or unequal relation, a cooperative or conflicting relation?) or the nature of society itself.

A particular example of the fallacies that individualist economics involves itself in, through its casual handling of the problems inherent in using subjective categories in attempts at scientific explanation, is that individualist economists will sometimes defend the reality of tastes by saying that they experience *themselves* as having tastes and making choices, so that they know that tastes exist objectively. This begs the scientific question, since what is required is not a private affirmation of belief, but a public demonstration that tastes are a real and powerful explanatory structure.

In passing I might suggest what seems to me to be the moral of this argument, which is that a scientific study of human society must begin by thinking about groups of people large enough to survive and reproduce themselves (just as the scientific study of ants might begin by viewing the anthill as the basic unit of analysis). In studying these groups it is sometimes necessary and always desirable to explain what subjective experience the people in the group will have of the material reality of their social existence, but we cannot expect to deduce much about social structure by starting from individual behavior.

II.

Let me now examine briefly the attempts to use individualist economic theory as a policy tool. The major concepts used in this endeavor are the idea of allocational efficiency and the idea of equity.

Economists often criticize and recommend policies from the point of view of allocational efficiency, that is, from a point of view that purports to represent the interests of the system as a whole. Antitrust policy, from this point of view, ought to be used to prevent deviations of price from marginal cost due to monopoly power. I gather that there is a good deal of frustration among antitrust economists because antitrust policy in reality does not systematically follow this goal. Some people are prosecuted even though their prices are close to marginal cost, and some whose prices are far above marginal cost are left alone. The same frustration reappears in most policy fields. Political decisions very often ignore the considerations of allocational efficiency.

Politics consists of struggles between groups of people. Things happen in politics because some particular group or coalition of groups becomes strong enough to enforce its own particular project on others. Allocational efficiency cannot be effective as a critical concept in influencing policy because it does not represent (in its own terms) the aims of any particular group. The idea that it could be effective arises because individualist economic theory represents government policy in an incorrect way, as the attempts of rational individuals to achieve collective ends, rather than as the tool which one class uses to enforce its desires on other classes.

The concept of equity as economists use it might at first glance seem to be a better way to approach policy problems. But I think it is also destined to be decoupled from politics. First, economists in talking about equity still try to make statements about properties of the whole society from outside it. Abstract considerations of equity are likely to move very few people politically. But, more important, classes in society struggle over their relations in a human and social sense: Who has power over whom in day-to-day life? When "equality" or "inequality" arise as forces in politics, they refer to equality as a characteristic of a social relation, and political struggles are most often undertaken to end a particular relational inequality. Unequal social relations (worker-capitalist, black-white, woman-man) always have the effect of producing unequal outcomes in the distribution of goods and services, but unequal outcomes are the symptom, not the disease itself. Equity in the sense that economists tend to use it errs in examining society at the level of distribution of goods and services alone. It ignores the relational inequalities that lie behind distributional inequalities, relational inequalities that I think can only be understood by looking through distribution to the process of production itself, and the unequal social relations that characterize production.

To sum up, then, the categories of allocational efficiency and equity are themselves the reason why individualist economic theory does not offer a reliable guide to or represent a powerful influence on real policies. The idea that it could or should have such an influence arises from an incorrect representation of social and political reality and government, one that ignores class differences and that understands government action as an attempt to achieve consensual ends rather than as always representing the particular interests of a victorious coalition.

III.

If individualist economic theory has serious problems as a scientific explanation, and if it also fails to reflect the real categories of political struggle and thus fails to influence political practice, why does it survive as a living body of ideas? I think that to answer this question we must turn our attention from the ideas themselves to the people who have them, that is, economists. (The argument of this section owes much to Louis Althusser's essay.)

In fact, we must address the question raised earlier: What is the relation between objective social reality and people's subjective experience of it? People generally have definite ways of perceiving social reality which is sometimes called ideology. More important than the particular content of people's view of the world are probably the categories by which they order their experience. A pertinent example is whether a person perceives human society as a unitary thing in itself, surviving through the medium of individuals, but fundamentally molding individual behavior (the anthill), or whether the person perceives individuals as the fundamental and natural unit and society as an aggregation of individuals (the ant). I argue that the way a particular group of people will image reality and their relation to it, its ideology, depends on the actual location of the group in the social division of labor, but not in a simple, direct way.

In considering individualist economic theory as the ideology of economists, it is important to think of ideology as a tree-like thing. Certain categories and understandings are common to everyone (or almost everyone) in a given society at a particular historical moment. This common system of ideas is sometimes called "dominant ideology." Individualism seems to be such a category in our society today. But there will be variants or elaborations of these common understandings which will be the peculiar preserve of certain groups that occupy a certain place. Individualist economic theory, insofar as it is individualist and rationalizes capitalist social relations in terms of individual interests, partakes of the dominant ideology and tends to reinforce the reproduction of capitalist and social relations.

But I think it is a mistake to understand the ideas that are specific to economists (for example, allocational efficiency) as direct apologies for capitalism. For one thing, individualist economic theory as an apology for capitalism, or as a mystified expression of the interests of the capitalist class, is hopelessly out of date. Very few people can bring themselves to believe in the competitive model as an accurate representation of reality for very long (say, longer than the time it takes to write an exam). As a guide to policy, individualist economics tends to come up with policy recommendations (deregulation, for instance) that are at variance with the expressed interests of dominant capital (and are therefore futile).

I suggest that individualist economic theory as a particular variant of the dominant individualist ideology continues to live because it represents the relation between social reality and the economist in a way that helps economists come to terms with their own fate. In addition to misrepresenting real social relations (representing unequal relations in production as equal relations of exchange, for instance) individualist economic theory creates an image of the economist as a technician standing outside social reality as a system, ready to intervene on behalf of the interest of the system as a whole. This image is imaginary and defective, because society is not a "system" in the sense of an object designed for some particular purpose, but a process of struggle. Thus there is no sense in which there is an interest of society as a whole, only the interests of particular classes or groups. The organization of society is not a technical question because institutions are not designed to perform certain functions; they are rather molded historically out of a process of struggle and compromise, so there is simply no place for a technician to intervene from outside to adjust or improve them.

Despite its defects as a representation of the situation, this image helps economists keep going in their real position, which is as a part of social reality and social struggle, not above it. The image distances the economist from what is going on. It is common for economists to view what is actually happening as a series of "mistakes" (that is, deviations from allocational efficiency or equity) and therefore in a sense not completely real (because things could have happened another way). The economist can take up a stance of humorous or sorrowful detachment toward what is happening.

The real place of economists in the social division of labor is as a part of the state apparatus (whether we are directly paid by the government or not) charged with collecting and ordering certain data about the economy, analyzing the data and trying to use it (usually without the help of any theory beyond an empirical faith in extrapolation) to predict consequences of policies, taking part in debates over state policy, and reproducing ourselves as a class of specialists. These activities are in most cases in support of the ruling coalition of class interests at any moment.

The practice of economists, I argue, is a practice of intervention on a particular side in political struggles. Individualist economic theory represents the economist as intervening, but not on any side—rather on behalf of the system as a whole. This contradiction is overcome in most cases by a logical leap, which identifies the real, actual government (if not the specific administration of the moment) with the interests of the whole society. This is a leap because government, like the market, money, corporations, and universities, grows out of historical struggle and has an objective reality that has nothing to do with the somewhat metaphysical notion of the interests of society as a whole. This leap is supported within individualist theory by models of government as a supplier of public goods or as an umpire in market failure situations—models which elaborate on the necessary leap without bridging the gap and which are themselves clearly defective representations of reality.

What is most striking to me about this situation is that economists in most cases enter the field critical of society and eager to understand its problems. We end in most cases by supporting the dominant class coalition and rationalizing or explaining away conflicts and contradictions in the society. Individualist economic theory has more than a static existence. It is a road of personal compromise that each human being who becomes an economist travels as social reality molds his or her practice. If we understood this more clearly about our own lives, we might not be so likely to accept misrepresentations of an individualist nature of other people's lives. People who become critical of social institutions and practice often sense that they need a critique or analysis of society as a whole to be successful in changing it. Individualist economic theory, which represents the society as a whole in mystified terms, is what you will be offered if you set out on that quest in our society.

I think that this is not a satisfactory state of affairs. The ineffectiveness of policy economics is a symptom of a deeper bind that economists are in. Frustration over the irrationality of regulatory policy or antitrust legislation or whatever is the way economists experience their own dependence on the dominant class coalition. That individualist theoretical categories like allocational efficiency and equity do not correspond to the effective categories of political struggle is a consequence of the ideology in which economists live and are able to function. It seems unlikely that a change in the content of economic theory can resolve this bind, which is personal and social as well as theoretical. I suggest instead that we must change our mode of response to social reality. Policy economics now consists of rhetorical, judgmental, "objective" written or verbal commentary on social problems. I think we must engage ourselves directly and consciously in the struggles of society on one side or another. Once our practice changes in this direction and we begin to experience ourselves as participants in social struggle, there will be a revolution in economic theory, but I cannot see how such a revolution can come first or by itself.

REFERENCE

L. Althusser, "Ideology and Ideological State Apparatuses," in *Lenin and Philosophy and Other Essays,* New York 1972.

Law and Economic Theory: An Economist's View

By Alvin K. Klevorick*

Taking stock of the realized and potential contributions of any area of intellectual endeavor is a difficult task. The assessment process is even more difficult if the field is relatively young and if it is interdisciplinary. Such is the case with law and economic theory. In this paper, I will attempt to give my view, as a participant in the law and economics enterprise, of the types of contributions economic theory can make to law—to legal decision making, to the study and development of legal doctrine, and to the study and analysis of legal structure. These views should be understood as simply *an* economist's view. I make no attempt to describe what must be a wide spectrum of views held by economists working in the area, and I certainly make no pretense to present the views of the lawyers engaged in this joint enterprise.

The first kind of contribution an economic theorist can make in law arises when economic concepts become important in understanding some aspect of a particular legal case. Although the overall question raised by the case may not be economic in nature, at some point an understanding of how markets work, how markets value commodities, services, and assets, and how individuals interact in their economic roles may become critical in deciding the ultimate disposition of the case. One can take an example as basic as the valuation of a capital asset. Suppose, for example, that through some set of circumstances (which we will not inquire into now), A has destroyed B's widget-making machine. The latter has no sentimental attachment to his machine, but values it only for the widgets it enables him to produce and sell. Without inquiring how the court reaches this position, suppose the court decides to award B an amount in damages equal to the value of the machine A has destroyed. An economist armed with his understanding of how markets value capital assets should be able to help the court in deciding the appropriate amount for B to receive in damages, given the court's decision that B should be compensated for the loss he has incurred because of A's destruction of the machine. An economist should also be able to help the court if B returns, as some real-world plaintiffs have, and asks for compensation, *in addition to* the amount the court has already awarded him, for the loss of profits he will incur until he can replace his widget-making machine with the new one he has just ordered.

In such a situation, the economist essentially plays the role of technician. He takes the problem facing the legal decision maker framed the way the legal decision maker has posed it, and he brings his expertise to bear in dealing with a specific part of the case. The economist draws upon his understanding of the way in which particular functions of society are performed, and he uses that understanding to shed light on specific issues in a given case. It is not a terribly imaginative role, but it is undoubtedly a terribly important one.

The second type of contribution an economist can make to law might be described as the economist in the role of "supertechnician." Once again, the economist takes the problem as set by the lawyer or by the legal decision maker, but in this instance the entire structure of the problem area has economic roots. The objectives and design of the institutions and doctrine are explicitly stated in economic terms, and the economist is called upon to evaluate and give advice about the best ways to achieve the specified objective(s). Most of the traditional areas of interaction between law and economics would fall, I think, into this category. They include, for example, the areas of antitrust law, public utility regulation, and labor law. When the issues at hand concern the most efficient way, or at least more efficient ways, of structuring an industry or a sector of the economy or the relationship between economic agents, the question comes ready-made for the economist to make an important contribution.

---

* Yale University.

One particularly important function the economist can play in his role as super-technician is to question whether the legal decision maker's general approach in a particular area actually helps to achieve the end the legal decision maker has in mind. For example, putting aside for the moment the multiplicity of reasons and rationales that might lie behind the institution of public utility regulation, suppose the legal decision maker asserts and believes that the primary objective of a system of public utility regulation is efficiency in the economist's strictly defined sense. The economist should be able to point out the kinds of characteristics commonly associated with natural monopoly and the kinds of characteristics which might suggest the need for special concern about the efficient functioning of an industry. He can help to focus the discussion to see whether the industry for which regulation has been proposed is indeed a natural monopoly. If the answer is yes, the economist can—and should—then suggest to the decision maker the variety of possible responses to the natural monopoly situation, responses ranging from having the natural monopoly privately owned and unregulated, to the general type of public utility regulation we have today, to an alternative type of regulation, or, finally, to public enterprise.

One thing the economist should do is provide the decision maker with an evaluation of the advantages and disadvantages of each of these alternative approaches. He should frame the discussion so that it encompasses not only the direct impacts each of these systems will have on the efficiency of the sector or industry whose regulation is being contemplated, but also the side effects or indirect effects that such regulation might have on this industry and its consumers and on other industries and other consumers in the economy. It would also seem to be incumbent upon the economist to try, to whatever extent possible, to indicate the income- (or wealth-) distributional impacts of such a regulatory policy. For although we have assumed (for the sake of this discussion) that the averred and true goal of the decision maker's use of regulation is to achieve efficiency, there are undoubtedly some tradeoffs between efficiency and distribution in the choice of a response to the natural monopoly problem, and the economist is particularly well suited to the task of pointing out what kinds of tradeoffs exist. Of course, while most of my discussion is concerned with the role of the economic theorist, the kinds of judgments that would be important, indeed critical, in such a balancing act are empirical judgments (and empirical guesses if hard data do not exist concerning these alternative costs and benefits). But even the theorist can help, and should help, to focus the discussion at the level of a comparative institutional analysis—a comparison of the alternative systems available for trying to achieve the given objectives.

In the context of this role as supertechnician, the economic theorist can also have a considerable input in designing the particular legal structure that is eventually chosen to cope with the given legal problem. For example, in the public utility area the economic theorist may be able to indicate which kinds of regulatory structures are most likely to achieve the goal of efficiency. He would be able to delineate, from an efficiency point of view, which kinds of issues are best handled using automatic adjustment clauses and which types of issues are best thrashed out in an alternative procedural setting.

In sum, when the economic theorist acts as a supertechnician for the lawyer, he draws on his expertise in analyzing a problem posed by the lawyer, just as he did in the role I described as economist as technician. But now the problem he is addressing is in its own terms economic in nature.

The third role I see for the economic theorist in the joint enterprise of law and economics is one that has come to flourish quite recently. Put briefly, it envisions the economist or economic theorist as the propounder of a new vocabulary, a new analytical structure for viewing a traditional legal problem. In contrast to the economist's approach in the first two categories of interaction I discussed, in this third role he no longer takes the problem as framed by the lawyer. Rather he takes the general problem area with which the lawyer is concerned—say, torts, or property, or procedure—and poses in his own terms—that is, in economic terms—the problem he sees the legal structure or legal doctrine confronting. He provides, thereby, a different way of looking at the legal issue which yields alternative explanations of how current law came to be what it is and new proposals for new law.

In the area of torts, for example, some economic theorists have posed the problem facing society as minimization of the expected social costs of accidents, taking account of: the costs incurred when accidents occur and the probabilities that such accidents will occur; the relationship between those accident costs and probabilities and the steps people take to avoid accidents; the costs of the resources devoted to avoiding accidents; and the costs of any administrative structure used to make decisions about the optimal level and allocation of these several kinds of costs among members of society. Economic theorists have used this type of framework to discuss and evaluate the traditional negligence rules, to provide a critique of the fault system of accident law, and to evaluate new proposals for automobile accident law, for example, no-fault plans. To take another example, the design of a system of civil procedure, criminal procedure, and judicial administration has been formulated by some lawyer-economists as the problem of minimizing the sum of the social costs of resources devoted to such procedural systems and administrative mechanisms plus the error costs of those procedures. This framework has then been used to evaluate various specific aspects of procedure and judicial administration: for example, the discovery rules of the Federal Rules of Civil Procedure, the bail system, the existence of plea bargaining, and the use of the jury.

This third role is probably the most exciting one the economic theorist can currently play. This form of interaction with the lawyer appears to be the most creative and the most challenging. Hence, it is interesting to ask whether particular kinds of problems confront the economist when he presents a new vocabulary or a new structure for analyzing a legal problem. I think two such problems are particularly important—one arises from a constraint imposed by the economist's tools while the other is a limitation imposed by the lawyer's formulation of the problem area.

The constraint imposed by the economist's tools is perhaps seen best in the context of a concrete example. Consider the problem of designing a set of liability rules for accident law in the world in which we live today. It should be clear that any attempt to analyze reality, to make meaningful statements about it, requires some abstraction from the richness of the phenomena being studied. The lawyer's world view is an abstraction just as the economist's is. The economic theorist's approach is to begin with the simplest possible model which captures important structural elements of the problem and to analyze that model as carefully as possible. As a theorist, I think this is a wise strategy. Having analyzed the simplest model, one might then proceed by relaxing some of the unrealistic assumptions which made one's first construction so tractable. For example, one might assume at the first step that all accident costs are, in fact, known to all parties making decisions about what levels of care to exercise. The second level of analysis might involve a relaxation of that assumption and a recognition that some parties will not be fully aware of these costs. The single (certain) values individuals attached to accident costs in the first model might now be replaced by subjective probability distributions people have about the accident costs that will be incurred.

Now suppose that in the simplest model several different liability rules would each yield the full-information social optimum by leading individual decision makers to attain the social-cost-minimizing combination of care levels. As one relaxes the assumptions of the model, the set of liability rules that will achieve the social optimum will undoubtedly become smaller. Further relaxations of the original assumptions—further gropings toward a more realistic representation of the world and people's decision-making processes—will undoubtedly reduce even further the number of liability rules that yield the social-cost-minimizing solution. Indeed, it is most likely that before long the set of socially optimal liability rule structures will become empty. In short, there is no first-best solution in the real world or, for that matter, in the most complicated, yet tractable model we may construct.

The question then is: If the economic theorist is going to use his theoretical analysis to make policy recommendations, how should he proceed? It would be tempting simply to consider those policies that were socially optimal in, say, the next-to-the-most complicated model he had. The problem with this, of course, is that the order in which the theorist relaxes his original assumptions to arrive at the most complicated model he analyzes—that is, the path that leads to the continually shrinking set of socially optimal policies—has no particular normative justification. He could follow an alternative complicating path which might well generate an alternative second-best policy.

The realization that we are making decisions in a second-best (indeed, an *n*th-best) setting does not imply that the theorist's models and analyses should be dispensed with or ignored. Nor should the second-best nature of the decision problem be used by the economist as an excuse for avoiding statements of a policy nature. To evaluate alternative policies, however, the theoretical model must be used in conjunction with careful, empirically rooted judgments (guesses?) about which of the theorist's original simplifying assumptions are most likely to affect the applicability of his model's results. And the order in which the theorist relaxes these assumptions in building successively more realistic models need not reflect the empirical importance of the various assumptions. Hence, there is no reason to evaluate second-best policies by simply removing, in reverse order, the complications introduced to obtain the most realistic model.

The theorist can contribute in two ways to the comparative institutional analysis the law and economics enterprise should produce in such situations. First, by providing a new and different way of looking at the traditional legal issue, he will, hopefully, provide clarification and new insights into that issue. Second, he will provide an indication of the kinds of empirical guesses that are needed in order to use his theoretical investigation for policy formulation in the world in which lawyers actually operate.

A different, but equally important, difficulty arises when the economic theorist provides a new vocabulary for a problem which the lawyer poses in a form not really amenable to economic analysis. While the theorist might be tempted to show how his analytical structures can be applied to particular issues in such areas, it is simply the case that his tools are inappropriate to the overall task. I have particularly in mind those areas in which lawyers or legal scholars have framed the legal question in process-oriented rather than outcome-oriented terms, that is, as a question about how a decision will be made rather than about what the decision will be. Questions of institutional competence, questions of whether a particular issue is one that should be taken up by legislatures or by courts, are questions which are not easily comprehended within the economist's framework.

The issue as posed by the lawyer in such cases does not seem to be rooted in the search for some neatly defined social optimum—be it the maximum of a social welfare function or the minimum of some social cost function. The problem is not framed as the search for an *efficient* way of reaching a decision or deriving a structure of rules, but rather as the determination of the "*appropriate*" way of reaching a decision or the "*appropriate*" institution to use to decide the particular issue. And the question of appropriateness is most likely to be resolved by appeals to history, to political theories of democracy, and to sociological theories of the role expectations of different members of the society. This is not to say that the economist's analytical structures cannot provide some help in examining these questions. In particular, when the lawyer appeals to political theories of democracy to help decide which institution is appropriate, he may well benefit from consideration of voting models and other models of the political process developed by theoretical economists and political scientists working in the area of public choice. The point simply is that when the lawyer or legal scholar has framed the problem in process terms, focusing on the appropriateness of one institution rather than another to perform a particular function or to make a particular decision, the lawyer is least likely to be receptive to the economist's vocabulary, rooted as that is in terms of a search for efficient structures, for efficient decision-making processes, and the like.

To be sure, questions of relative institutional competence can be stated in economic terms. One can talk about which institution has a comparative advantage in performing which function. We can ask whether courts or legislatures are "better" at developing liability rules, whether courts or administrative agencies are "better" at determining the value of a fair rate of return to a public utility, whether legislatures or courts are "better situated" to decide which interests are fundamental and which interests are not (alternatively, which wants are merit wants), or whether the judiciary or the legislature or the executive is the appropriate institution to decide if and when prior restraint over publication should be exercised. One can use the terminology of economics and talk about the costs and benefits of using one institution rather than the other. But when the legal scholar has posed the problem in process terms, the analytical structure the economist offers for the given legal issue may not prove very helpful to the lawyer.

Let me try to illustrate this point with two examples. The first concerns freedom of speech. I believe that the economist (or the lawyer-economist) can contribute to our understanding of and deliberation about the appropriate scope for freedom of speech by taking seriously a concept which many Justices and legal scholars have invoked in their discussions of First Amendment freedom of speech issues: the marketplace of ideas. The economist can contribute by taking such references to the marketplace of ideas as something more than simply rhetoric—as, instead, an attempt to invoke what appears to the lawyer to be an appropriate analogy. Then, by bringing to bear his understanding of the ways in which markets work, of the potential sources of market failure, and of the possible responses to such market failures—all within a comparative institutional analysis framework—the economist may provide interesting observations (at least in efficiency terms) about possible arguments justifying and opposing limitations on freedom of speech in specific situations.

Exploiting the analogy of the marketplace of ideas, the economist can, I believe, provide a clarifying view of some aspects of the development of legal doctrine in the First Amendment area, for example, "the clear and present danger" test and its development and application in the case law. But suppose the legal scholar's formulation of the First Amendment issue takes the form of the question: Which institution or which branch of government is the *appropriate* one to determine the scope of First Amendment rights? For example, suppose that the majority of a duly elected legislature believes that advocacy of a particular political philosophy poses an imminent danger to the nation. The legislature enacts a statute making advocacy of the particular political views a criminal act. The legal scholar might ask: How deferential should a court be to the legislature's judgment if a suit is brought challenging the law? The economist's analysis and interpretation of the development of judicial doctrine may not be very helpful to the lawyer or legal scholar in answering this question.

The second example concerns whether juries in criminal trials should be representative, whether they should represent a cross section of the community. Here I think the economist can help in evaluating how effective representativeness of the jury is in furthering performance of the jury's fact-finding function. He can suggest and explicate the analogy between the selection of a jury and the selection of a portfolio of assets by an investor. Pursuing this analogy, with the consequent delineation of the similarity between representativeness of a jury and diversification of an investment portfolio, the economist can draw upon portfolio selection theory to suggest the kinds of circumstances under which representativeness would make the jury a more effective fact-finding body and the types of situations in which representativeness would not serve that end. I think that introduction of the economist's vocabulary in this discussion is interesting and helpful. But if the question of jury representativeness is framed not in the instrumental sense of achieving some further objective, like fact finding, but rather in more process-oriented terms, which would have a representative jury as the *appropriate* type to decide criminal cases, then the economist's vocabulary is much less likely to clarify and illuminate the issues for the lawyer.

We come now to the question of what happens after the economist has proposed a new vocabulary for viewing a traditional area of legal concern. One possibility, of course, is that the new vocabulary will simply be rejected, and I have tried to indicate at least one reason why that may occur. A second possibility is that the new vocabulary, the new way of viewing the problem, will come to displace the old. While I regard this outcome as highly unlikely, it nevertheless suggests the major danger that I see, and that I hope will be averted, in the application of economic theory to legal problems: namely, a type of intellectual imperialism on the part of the economic theorist or the lawyer-economist. It is the danger that, having created or at least suggested a new vocabulary for viewing the legal issue, the economic theorist or the lawyer-economist will develop—or even worse, try to generate in others—a form of tunnel vision in which he sees the new vocabulary as the only "reasonable" one for viewing the area. While I firmly believe that economics can make a contribution to law and that economic theory in particular can make such a contribution, I do not believe that the analytical grid the economic theorist imposes on the legal problem is the only one worth examining.

The most likely result of the economist's suggestion of a new analytical structure is the coexistence of his new approach to the problem with other, perhaps more traditional, approaches to it. For example, thinking of civil procedure as designed to minimize the sum of the direct costs and error costs of procedure will come to coexist with an understanding of certain elements of procedure as generating or guaranteeing a sense of fairness in the way our legal system operates. Our system of tort law will be understood both in terms of minimization of the social costs of accidents and in terms of noneconomically based ideas of causation and reciprocity.

Economic theory's contribution to law is likely to be greatest if, in proposing new analytical structures for traditional legal issues, the economist or lawyer-economist takes a constructive eclectic view and welcomes the contributions other social scientists may be able to make to our understanding of these areas. For psychologists, historians, sociologists, and others can also suggest new and potentially useful approaches to legal issues, approaches which will undoubtedly complement the economist's with its relative inattention to a variety of questions—like historical development, the formation of individual and social tastes, and so on. The economic theorist's new formulation essentially provides a new metaphor for viewing the particular area of legal concern. But the economist's suggested metaphor is only one among many, and in attempting to understand, to analyze, and to aid in the development of the law, an approach which draws upon the variety of ways the several social sciences analyze human behavior is likely to be the most helpful and the most successful.

SOURCES AND USES OF PANELS OF MICRODATA

The National Longitudinal Surveys: New Vistas for Labor Market Research

By HERBERT S. PARNES*

For nearly a decade the Ohio State University Center for Human Resource Research and the U.S. Bureau of the Census, under separate contracts with the U.S. Department of Labor, have been engaged in the National Longitudinal Surveys (*NLS*) of labor market experience. Four subsets of the United States civilian population are being studied: middle-aged men who at the inception of the study were 45 to 59 years of age; women 30 to 44 years old; and young men ("boys") and young women ("girls") between the ages of 14 and 24. A rich data bank is being accumulated that has already served as the basis for thirteen comprehensive research monographs and over seventy-five specialized studies completed by staff members of the Center for Human Resource Research and other researchers throughout the country.[1] The purpose of this paper is to describe the nature and availability of the *NLS* data and to illustrate the more or less unique kinds of analysis they make possible.

I. The *NLS* Data

The cohorts comprising the *NLS* were chosen because each is confronted by special labor market problems. In the case of the men, for example, skill obsolescence, deteriorating health, declining vigor, or age discrimination may lead to premature withdrawal from the labor force or other manifestations of labor market difficulty. The special problems of the women are those attending reentry to the labor force after periods of absence during which there may have been a deterioration in both vocational and labor market skills. The two groups of youth face the difficult process of occupational choice and the period of accommodation to the labor market during which turnover and unemployment rates are high.

For each of these four population groups, a national probability sample of approximately 5,000 individuals was drawn by the Census Bureau by a method analogous to that used in the Current Population Survey (*CPS*). However, in order to permit statistically reliable estimates for blacks, the sampling ratio in predominantly black neighborhoods was three to four times as high as in white areas. As a result, each sample contains roughly 3,500 whites and 1,500 blacks.

A very useful aspect of the sampling design is that many of the households yielded more than one respondent, frequently across cohorts, and it is possible to relate to any given respondent other respondents who were selected from the same household. To illustrate, there are 567 households with at least one subject in the men's sample and one in the women's, most of these representing husband-wife combinations. Similarly, there are about 900 man-boy and 800 man-girl combinations; 1,400 woman-boy and 500 woman-girl combinations; 1,800 boy-girl combinations; and in excess of 300 households in which man, woman, and at least one youth are included in the *NLS* samples.

---

* Professor of economics and research associate, Center for Human Resource Research, The Ohio State University. I am grateful to my colleagues Carol Jusenius, John Grasso, Gilbert Nestel, and Steven Sandell for useful comments on a preliminary draft.

[1] For a complete bibliography, see the *National Longitudinal Surveys Handbook*, beginning on p. 41.

Members of each sample have been interviewed periodically by *CPS* enumerators, using schedules prepared by the Center for Human Resource Research. Originally planned to cover a five-year period, the surveys have been so successful and attrition so small that they have been continued beyond the originally planned expiration dates. As of the end of 1974, the men had been surveyed in 1966, 1967, 1968 (mail), 1969, 1971, and 1973 (telephone); the women in 1967, 1968 (mail), 1969, 1971, 1972, and 1974 (telephone); the boys annually between 1966 and 1971 and by telephone in 1973; and the girls annually between 1968 and 1973. As of the most recent interviews, attrition had claimed only 21 percent of the original sample of men, 15 percent of the women, 23 percent of the boys, and 14 percent of the girls. In the case of the men somewhat over half of the losses were attributable to death. Among the boys, induction into military service accounted for much of the attrition between 1966 and 1971, but respondents reentered the sample as they returned from service; the number of completed interviews was slightly greater in 1973 than in 1971.

In a brief paper it is possible only to hint at the variety of data that have been collected over the years of the *NLS*.[2] In most general terms it may be said that information has been obtained on virtually every aspect of labor market experience—in abbreviated form for that portion of the individual's career prior to the initial survey and in considerable detail for the period covered by the surveys. In addition, information has also been accumulated about factors that are hypothesized to influence labor market behavior and experience, e.g., health, education and training, family structure, a variety of attitudes, and several characteristics of the local area in which the respondent resides. Many of the questions included in the interview schedules are common to all four cohorts, e.g., those relating to current labor force and employment status, employment experience during the period since the previous survey, health condition, training experience, job attitudes, and sources of income and wealth. On the other hand, some are unique to one or two cohorts, e.g., retirement expectations and attitudes for the men; detailed educational histories for the boys and girls, including mental ability test scores obtained from the high schools they attended; and child care arrangements and preferences for the women.

II. The *NLS* Public-Use Files

Publication of the initial reports on the *NLS* generated numerous inquiries about the availability of the data, in response to which the Census Bureau agreed to prepare public-use versions of the data tapes very similar to those which they were preparing for the Center for Human Resource Research. These tapes had a number of undesirable characteristics—e.g., alphabetic codes—that made them difficult to use for many types of analysis. Thus, all purchasers had to perform "cleansing" operations that were more or less identical to those undertaken by the Center to make the tapes operational. In order to eliminate this duplication of effort, the Center agreed in 1973 to assume the responsibility for revising those tapes already released by the Census Bureau and for preparing all subsequent public-use tapes.

---

[2] A detailed description is contained in the *National Longitudinal Surveys Handbook*.

By the end of 1974, cumulated data files through the 1971 surveys for the men and women, the 1970 survey of the girls, and the 1969 survey of the boys had been released by the Center, each at a price of $300. It is contemplated that an updated file containing the data from each additional survey will be ready for release approximately one year after the Center receives the relevant tape from Census. On this basis, cumulative records through 1973 for the men, 1970 for the boys, 1971 for the girls, and 1972 for the women are expected to be forthcoming in 1975, each at the $300 price mentioned above. This price includes the computer tape(s) in any one of twelve forms desired by the purchaser, two copies of the documentation, and for those who wish it, a copy of the documentation on computer tape.

The documentation that accompanies the *NLS* data files is designed to facilitate access by users to a large and complex body of data, e.g., 2,540 variables in the case of the men's surveys through 1971. A very useful component is the *KWIC* (key word in context), an alphabetic index that provides a convenient means of identifying variables and locating them in the code book. For each variable, the code book displays a frequency count and indicates the interview schedule item number(s) from which the variable was derived.

As of the end of 1974, 69 data files had been ordered by 37 researchers or research groups throughout the United States. However, these figures understate the extent of utilization of the *NLS* data, since the Census Bureau had sold over 125 individual survey tapes to a total of 28 researchers prior to the revision of the files by the Center. Moreover, an unknown number of "bootleg" tapes are in use, having been copied from those purchased either from the Census Bureau or the Center. Under its contract with the Department of Labor, the Center serves as a central clearinghouse for research utilizing the *NLS* data. Since the spring of 1974 it has published a quarterly *Newsletter* that catalogues recent research reports and research in progress based on *NLS* data. Also included is information concerning release dates of future data files and errors that may have been uncovered in those already issued. The *Newsletter* is mailed to all known users of the data files and to other members of the research community who request it.[3]

III. Research Potential

Although a substantial portion of the data from the *NLS* has not yet become available for analysis, enough work has already been done at least to illustrate the potential contribution of the project to a better understanding of the operation of labor markets and of factors affecting the labor market experience and behavior of individuals. The richness of the *NLS* for labor market research stems largely from the longitudinal character of the data, from the inclusion of a variety of social-psychological variables along with detailed work histories, and from the existence of multiple-respondent households.

Perhaps the single most important contribution of longitudinal data is that they facilitate the identification of causal relationships that cannot confidently be identified in any other way. Take, for example, the relationship between attitudes and behavior. In cross-sectional data such relationships are ambiguous, since one cannot be certain whether the attitude produces or reflects the behavior. Does job dissatisfaction lead to turnover, or does an association between the variables simply mean that individuals who quit jobs are likely to rationalize their behavior by reporting (retrospectively) that they were unhappy? When attitudes measured at one point in time can be related to *subsequent* behavior, such ambiguity disappears. The *NLS* data for middle-aged men have clearly demonstrated that the degree of job satisfaction *predicts* the likelihood of a voluntary job separation and that a commitment to work in general, as well as satisfaction with one's particular job, decreases the likelihood of early retirement (Parnes et al., chs. 3 and 5).

---

[3] Requests should be addressed to Center for Human Resource Research, 215 W. 11th Avenue, Columbus, Ohio, 43201.

To cite another example, Thomas Mahoney's analysis of cross-sectional data found a relationship between the attitudes of married women toward the "proper" role of wife and mother and their labor force participation, but G. Cain (pp. 39–40) has criticized such evidence for its circularity. The ambiguity would appear to be removed by the *NLS* data shown in Table 1, which relates the attitudes expressed by young women *before their marriage* to their labor force activity after marrying and bearing children. Those who had expressed the most permissive attitudes toward labor market activity by young mothers in 1968 were, *ceteris paribus*, significantly more likely to be in the labor force three years later than those with less favorable attitudes.

TABLE 1—ADJUSTED LABOR FORCE PARTICIPATION RATES OF MOTHERS 17–27 YEARS OF AGE, BY PREMARITAL ATTITUDE TOWARD EMPLOYMENT OF MOTHERS OF YOUNG CHILDREN, 1971

Attitude[a]	Number of Sample Cases	Adjusted Labor Force Participation Rate[b]	
Opposed	104	33.5	
Ambivalent	123	31.8	
Favorable	150	48.7	
Total[c] or average	381	38.8	
F-ratio			3.59*

[a] Based on scores on a series of questions exploring respondent's reaction to propriety of work by women with preschool-age children under alternative specified circumstances.

[b] Adjusted labor force participation rates obtained from multiple classification analysis with three explanatory variables in addition to attitude: race (white, black); education (0–11 years, 12 years, 13+ years); husband's income (under $5,000; $5,000–$7,999; $8,000 or more).

[c] Total includes four cases for which attitude was not ascertained.

* Significant at $\alpha \leq .05$.

Source: National Longitudinal Surveys.

Social psychological variables in the *NLS* have also been shown to exert an influence on earnings or occupational status independent of conventional human capital variables. For example, Paul Andrisani and Gilbert Nestel report that a belief in the efficacy of individual initiative contributes to the level of average hourly earnings of middle-aged men (Parnes et al., ch. 6). As another illustration, Parnes and Andrew Kohen have shown that young men who manifested superior occupational information in 1966 enjoyed greater earnings and higher occupational status two years later than their more poorly informed counterparts.

The usefulness of longitudinal data in clarifying causal relationships is, of course, not confined to instances in which one of the variables is attitudinal. For example, finding that the receipt of training by middle-aged men between 1966 and 1971 was associated with a net earnings advantage in 1971 (controlling for such other factors as education, health, and region of residence) Arvil Adams went on to demonstrate that the trainees-to-be had already enjoyed higher earnings in 1966 (again controlling for the same variables) (Parnes et al., ch. 2). Thus, training was found to be a selective process, presumably attracting the more highly motivated or otherwise more productive individuals. To put the matter differently, some part of what would doubtless have been identified by a cross-sectional analysis as training's contribution to earnings was found to have reflected an incompletely specified model—i.e., the failure to control adequately for factors associated both with earnings and the probability of receiving training.

Quite aside from its longitudinal character, the *NLS* data provide a particularly rich data source for exploring the extent and nature of labor market discrimination. As is well-known, the principal difficulty in this type of research lies in being able to control for all factors that affect productivity. The social-psychological variables that exist in the *NLS*, along with measures of human capital and detailed records of labor market experience, provide a better basis for doing this than most other bodies of microdata. The question of racial discrimination can be pursued within each of the cohorts, while the issue of sex discrimination can be studied on the basis of the data for the girls and boys. A limited amount of research on racial discrimination based on *NLS* data has already been reported by Kohen, Robert Flanagan, and Robert Hall and Richard Kasten.

As has been indicated above, the existence of substantial numbers of multiple-respondent households in the *NLS* affords unusual research opportunities stemming from the fact that detailed information has been obtained directly from each individual relating to personal characteristics, to labor market behavior and plans, and most importantly to attitudes. One use of such data is in methodological studies of the consistency of response to factual questions by different members of the same family (e.g., Michael Borus and Nestel). Second, a number of fascinating questions may be raised concerning the consistency of various types of attitudes among family members, e.g., the relationship between the educational aspirations of a young man and the aspirations for him expressed by his father. Third, the extent to which the attitudes and/or labor market behavior of one family member affect the current or subsequent characteristics or behavior of other members of the family may be explored. An interesting illustration is Belton Fleisher's study of the effect of a mother's allocation of time between home and market production on the human capital of her son(s). Most models of labor supply assume that the behavior of individuals is guided by a desire to maximize the utility of the entire household, but the absence of data has generally prevented empirical research from making such a model operational. The multiple-respondent household in the *NLS* should go a long way toward eliminating this deficiency, albeit with a relatively small and perhaps nonrepresentative sample.

Another category of research that may be served by the *NLS* data must at least be mentioned—the study of gross changes in a variety of educational and labor market variables—e.g., enrollment status, labor force and employment status, occupational and industrial affiliation, etc. A very important line of investigation subsumed in this category relates to changes in income. A beginning has already been made in the analysis of the extent and sources of income instability among the men and boys (Kohen, Parnes, and Shea), but this will doubtless be expanded to additional cohorts and over longer periods of time as the studies continue.

Finally, an important contribution of the *NLS* stems from the opportunities they afford for the replication of studies. The availability of identical data for a number of years allows a researcher to examine the consistency over time of relationships or estimates derived from data for a single year. Also, the widespread distribution of the public-use tapes permits researchers to attempt to replicate the findings of complex research designs and methods of analysis employed by their colleagues. Both of these processes should enhance the confidence with which research findings may be accepted.

IV. Conclusion

The National Longitudinal Surveys permit a more effective investigation of a number of types of questions relating to the operation of labor markets than has hitherto been possible. The types of research that have already been based on the data illustrate the potential, and a good portion of future research will doubtless be similar to that already executed, but will have the advantage of additional waves of data as they become available. If, as now appears likely, the surveys are continued to cover a full decade, additional types of research will be possible. For example, by 1976 a third of the sample of men will be at or above the conventional retirement age of 65 and an even larger proportion may have actually retired, which will permit an analysis of the relationship between retirement expectations and actual retirement behavior. Similarly, by 1976 and 1978, respectively, the substantial proportions of the boys (60 percent) and girls (50 percent) who were enrolled in school at the time of the initial survey will, with few exceptions, have left school, permitting comprehensive analyses of the process of adjustment to the labor market. In any case, the *NLS* data base will almost certainly provide a fertile field for labor market research for many years to come.

REFERENCES

M. Borus and G. Nestel, "Response Bias in Reports of Father's Education and Socioeconomic Status," *J. Amer. Statist. Ass.*, Dec. 1973, *68*, 816–20.

G. Cain, *Married Women in the Labor Force: An Economic Analysis*, Chicago 1966.

B. Fleisher, "Mothers' Home Time and the Production of Child Quality," unpublished manuscript.

R. Flanagan, "Labor Force Experience, Job Turnover, and Racial Wage Differentials," *Rev. Econ. Statist.*, Nov. 1974, *56*.

A. Kohen, "Determinants of Labor Market Success among Young Men: Race, Ability, Quantity and Quality of Schooling," unpublished doctoral dissertation, Ohio State Univ. 1972.

———, H. Parnes, and J. Shea, "Income Instability among Young and Middle-Aged Men," in J. Smith, ed., *The Personal Distribution of Income and Wealth*, New York 1975.

R. Hall and R. Kasten, "The Relative Occupational Success of Blacks and Whites," *Brookings Papers on Economic Activity*, 1973, *3*, 781–95.

T. Mahoney, "Factors Determining the Labor Force Participation of Married Women," *Ind. Labor Relat. Rev.*, July 1961, *14*, 563–77.

H. Parnes and A. Kohen, "Occupational Information and Labor Market Status: The Case of Young Men," *J. Human Resources*, Jan. 1975, *10*, 44–55.

———, A. Adams, P. Andrisani, A. Kohen, and G. Nestel, *The Pre-Retirement Years: Five Years in the Work Lives of Middle-Aged Men*, Columbus 1974.

Center for Human Resource Research, Ohio State Univ., *Nat. Longitudinal Surveys Handbook*, Columbus 1974.

Using Survey Data from the University of Michigan's Survey Research Center

By James N. Morgan*

There is a great deal of useful analysis that can be done with available survey data. The quality and quantity of such data are growing, as is the ease of working with it on computers. I want to talk first about some exciting analysis possibilities and data sets; second, a little about the process of locating and getting such data, particularly ours; and finally, about some of the statistical and methodological considerations in analyzing such data.

You might think that the original collectors of the data would already have published the most important findings. Not so. Much remains to be done and I shall mention some main approaches, then some of our more exciting data sets.

Sometimes it is useful to show that *assumptions* which lie behind a theory, or those that are used to test it, are incorrect. Use of self-employment as an indicator of a large transitory component of income is discouraged, for example, when we discover that blue-collar workers have much less stable incomes than the self-employed.

Many theories deal with individuals at the margin, but probability samples contain many people who really have no choices to make. We may need to do some careful selecting of subgroups that can be expected to respond to the particular economic variable of interest. In the labor supply area, for example, it is a fact that "Of all employed male heads of households in 1971 *less than 13 percent* held jobs in which they were free to vary their work hours at well-defined marginal wage rates and nearly half of those gained their flexibility of work hours by taking a second job" (Jonathan Dickinson). Studies of labor supply cannot ignore such constraints.

A logical next step is to search for the optimal disaggregation, that is, for differences in response parameters among subgroups of the population.

Another important kind of analysis deals with the impact of actual or proposed changes on different parts of the population. It is now a common question asked of any proposed changes: "Will it hurt the poor?" We recently pointed out that miles of commuting by car is relatively income elastic, so that a tax on gasoline would *in general* not be regressive. But we could also estimate how many exceptions there were to this rule that might call for special provisions to avoid hardship (Morgan, 1974).

One can develop a better set of assumptions on which to build economic models. To illustrate this: the recent migrants to urban areas, other things considered, earn less than those who were already there when they arrived. Studies of the payoff to migration must clearly be carefully done. Second example: the closer one lives to the center, and the larger the city, the longer it takes, on the average, to get to work. The reason is that not everyone works in the center of town and travel speeds drop dramatically as one gets closer. People who move don't minimize travel plus housing costs. They often move to places farther from where they work and at no savings in housing cost.

What does affect behavior, and why—misinformation, missing information, complex goals, or failure to understand economics? Optimal rules for information search, or lifetime allocation of time and consumption, or optimal inventories are extremely complex, hardly soluble by most of us. Long ago it was discovered that some businessmen kept inventories in amounts proportional to sales even though it was clear to economists that they should be proportional to the square root of sales!

---

* University of Michigan.

Most of the interesting questions of economic behavior are dynamic, yet much survey data is from single cross sections. In spite of this, a great deal of dynamic behavioral analysis is possible even from such surveys. Age cohorts reflect in some crude way life-cycle effects. Some prior information is available from memory. And when there are repeated national probability samples, one can identify subgroups on characteristics that do not change and compare them at different times. Such subgroup differences, which can be expressed as regression coefficients of dummy variables, can be used to separate differences that are effects of different behavior (coefficients) from those resulting from different proportions of the population in the relevant subgroups. Although the third set of components, involving correlations of those pairs of differences, are unavailable without reinterviews, they are likely to be small. There has also been some use of long series of annual or quarterly surveys in time series models directly. For a fuller discussion, see Morgan 1972, Herbert S. Parnes, and Thomas K. Glennan, Jr.

Indeed, one of the advantages of cross-sectional surveys over panel studies is that, being less expensive, they can use larger samples, and with larger samples one has more chance of isolating the marginal subgroups with choices to make or of policy interest.

As more and more data become available from studies which reinterview panels of respondents, data on change can rely less on memory, and the full advantage of the microdata in studying change is available. The models become more complex, and the problem of relating them to aggregate dynamic models arises.

Probably the most exciting and promising possibilities are in the simulation models that use if-then data and iterative computer simulation of the system.

The Survey Research Center has been conducting surveys and making its data available for more than twenty-five years. Let me mention *some* of our available data sets and try to say just a little about some of the possibilities they offer for further analysis.

I. Some Examples of Useful Data Sets

The data available from the Survey Research Center of the Institute for Social Research at the University of Michigan are generally from personal or telephone interviews with small national probability samples varying from 1,200 to 5,000. Such small samples are easy to handle though not ideal for dealing with problems relevant for very small subparts of the population. Sometimes there is oversampling of one end or the other of the income distribution, or of an age group, or of victims of auto accidents or work injuries. The data also contain a useful mix of quantitative economic data, along with attitudes, expectations, knowledge levels, insights, reasons, and demographic facts and background.

First are the annual Surveys of Consumer Finances, 1947–1970, with enough consistency of content, sample design, and methodology to provide trends on many things plus some built-in reinterviews. The data appeared in the *Federal Reserve Bulletins* until 1960, then in annual volumes from the Survey Research Center. There have been smaller surveys, several a year, containing the material on purchase plans, major purchases, and consumer attitudes that were also in the financial surveys.

These two sets of data have already been used in some time series models. There was a four-wave reinterview study of consumer debt use partially imbedded in the last few Surveys of Consumer Finances (see Gary Hendricks and Kenwood Youmans). A series of studies of urban residential choice and the journey to work was summarized in a book by John B. Lansing and Hendricks, and there was a study by Eva Mueller of what machine technology was doing to the worker. A study of early retirement and retirement plans combined a national sample with a sample of *UAW* workers close to retirement age. (See R. E. Barfield and Morgan; and Barfield.) A major study, oversampling lower incomes, of why there were still poor people in 1960 and what they were doing about their children, was followed by another which focused on the paid and unpaid work and related attitudes of families. (See Morgan, M. David, W. Cohen, and H. Brazer; and Morgan, I. Sirageldin, and N. Baerwaldt.) Special studies focused on planned residential environments, the chain of people who moved as a result of new dwellings, on auto accident victims in Michigan, and on the affluent and their work and investment behavior. (See Lansing, Robert Marans, and Robert Zehner; Lansing, C. W. Clifton and Morgan; Alfred Conard et al.; and R. Barlow, Brazer and Morgan.)

The richest, and still growing, data set comes from a national panel study oversampling lower income families, originally funded by the Office of Economic Opportunity and now funded by the Department of Health, Education and Welfare. Two volumes reporting the first five waves are out, a third including the sixth wave is in press, preliminary results from the seventh wave are being tabulated, and plans for the eighth wave are being made. (See Morgan et al., 1974.) Not only does this panel study contain income and work data, and more than usual background material, but for the first five years there were measurements of attitudes and behavior patterns thought to be related to economic change. The fifth-year interview contained a sentence completion test and an attempt to measure achievement motivation. The sample consists of all the individuals in the original sample of families, so those that split off from original families are followed, allowing studies of divorced people, two generations, etc. There is both a family data file and an individual data file, the latter containing data about each individual and about the family he/she was in each year. We have added more than usual environmental data from outside the interview, including county data on unemployment incomes, population growth, and area data on prices. Nonmoney income from assets (imputed rent) and from work are asked about, as well as subsidized food or rent. Hours of work, housework, commuting, illness, and unemployment are measured so that estimates of leisure time are possible. And an extensive set of background information on both husband and wife is available. Repeated measurements of the same set of attitudes and behavior patterns for five years allow studies of changes.

Social scientists in several universities are already analyzing the data and we summarize their work as well as our own in annual volumes of findings. The seven-wave family and individual data tapes will soon be available.

Two studies currently in progress that will probably be available by the end of 1975 should be of great interest. One is a national study of revenue sharing, interviewing administrators and elected officials on the effects of revenue sharing as they see it. The other involves two national samples, one with some oversampling of upper incomes and the other heavily loaded with very high incomes, asking about philanthropic giving of both time and money, related attitudes, and expected responses to further possible changes in the tax laws.

I should mention that this panel of speakers does not cover certain other important data collections or sets of data collections. The Office of Research and Statistics of the Social Security Administration has just issued a *Work Plan* for fiscal years 1975 and 1976 which describes a variety of things, including its continuous work history sample and a ten-year, longitudinal study of retirement, using interviews every two years of the same people. The massive Bureau of Labor Statistics expenditure survey with its five quarterly interviews plus diaries is probably known to most of you.

II. Locating and Getting Survey Data

A long-standing policy of the Institute for Social Research is that as soon as the first publication of findings from any survey is available the basic data are to be available to others for further analysis. The simplest procedure is to read our published reports to find a survey that has the data you want, then get the data and documentation. (See References.) The data library of the Economic Behavior Program (which handles all of the abovementioned studies) tries to work on a no-gain-no-loss basis which in practice has meant a charge of about $350 for the documentation and data from any major survey. For the large Panel Study of Income Dynamics, there are documentation volumes available as an intermediate step in securing the data.

Many studies oversample groups of special interest, such as blacks or low-income families. In order to have unbiased estimates of population parameters it is necessary to weigh each case by the inverse of its sampling fraction. Frequently, the weights supplied also adjust for differential response rates. It is commonly said that with a properly specified model such weights are unnecessary, but in fact our models are almost never properly specified or at least we do not know that they are. For instance, in a sample oversampling blacks, an unweighted estimate of the income elasticity of home ownership that failed to allow for an interaction effect with race (not just a race effect) would produce a weighted average elasticity that is too high since blacks have a higher income elasticity. Analysts usually do not know about all such interaction effects in advance. Many computer programs handle weighted data, and also carry along unweighted numbers for purposes of statistical inference.

Most samples are multistage, stratified, clustered probability samples, not simple random samples, and this produces "design effects"—departures of sampling variances from simple random. Since the losses in information per interview from clustering are usually greater than the gains from stratification, statistical inferences that ignore the design effect will exaggerate levels of significance and underestimate sampling variances. Most studies provide some indication of the average level of design effects. Efficiency losses per case tend to be smaller for more complex statistics like regression coefficients than for means. For a fuller treatment of survey methods and analysis you might want to read our survey methods book (Lansing and Morgan).

III. Some Statistical and Methodological Considerations

Economists trained in the statistical analysis of time series and the traditional econometrics will find that the shift to survey data changes the nature of the statistical problems. There are many more variables, often only loosely related to the theoretical concepts. There are many more "degrees of freedom" from the large sample sizes, so that we have a problem of choosing from among many possible (probably misspecified) models. To make matters more complicated, there are clearly measurement errors of various sizes in the explanatory variables, raising statistical problems that had all but disappeared from the econometric literature. The comparative assessment of the importance of different explanations (explanatory variables) requires some estimate or guess of their error variances. Recent controversies over the relative importance of genetics, parental background, and quality of education in explaining achievement or income have foundered over this issue. It is escape from the degrees of freedom problem provided by the thousands of cases available that changes things most, however. Almost any effect big enough to matter will be statistically significant, and it is the relative reduction in error variance that matters.

Without attempting a lecture in statistics, let me outline some of the problems and some of the suggested solutions: the dependent variable may be skewed, or have heterogeneous variance, requiring transformation or weighted regression (weighting by the inverse of its estimated variance). It may be dichotomous with problems of heteroscedasticity (if it is near 0, or 1) and of predictions <0 or >1. Some argue for probit or logit estimation and we have a multivariate logit program operating. But with reasonably good models, dummy variable predictors, and substantial sample sizes, we think such methods may be unnecessary. It may be a set of classes, the probabilities adding to 1. We have a computer search program that identifies subgroups that differ in their distributions, and another that treats membership in each subclass as a dichotomous (dummy) variable in a separate regression. Or one can convert the problem to a sequential set of conditional probabilities or use a multiple discriminant function. Finally, one can make use of the interrelationships, à la Arnold Zellner. Where there are dependent dichotomies that do not add but are related, such as use of a set of modern practices, they can be handled by methods proposed by M. Nerlove and S. J. Press. One may have a *set* of dependent *variables*. If they are shares that add to 100 percent, we have a search program that looks for subgroups with different share distributions, for instance of the fractions of income from earnings, transfers, or capital. If they are related dependent variables, there is a search program reported for this problem (see M. W. Gillo and M. W. Shelly). For estimation of one optimal model, the Zellner approach is presumably still appropriate. Canonical correlations seem to me to make too many restrictive assumptions.

The explanatory variables may be nonquantitative or skewed quantities. In either case conversion to a set of dichotomous (dummy) variables is usually simplest. (See F. M. Andrews, Morgan, J. A. Sonquist, and Laura Klem; and for a general search program to locate the best linear model, see Sonquist, E. L. Baker and Morgan.) They may be measured with varying errors, and although instrumental variables procedures are often suggested, appropriate ones are hard to find. One can think of the same problem as one of "unobservable variables" or variables only loosely related to the theoretical constructs. In overdetermined systems, these variables can be estimated along with some needed error variances by methods suggested by Goldberger and Jöreskog and available in a program called *LISREL*. (See especially chs. 1, 4, and 5 of Arthur S. Goldberger and Otis D. Duncan.)

There may be interaction effects, the impact of one factor depending on the level of another. Since there are so many possibilities, and they may not be appropriately handled with simple cross-product terms, our *AID-III* search program is often useful to narrow down the candidates. There may be causal paths among the predictors, as when race affects education which affects occupation which affects income. If they all go in one direction, the resulting recursive model is relatively easy to deal with, provided one does not have problems of scaling or interaction effects.

When there are reinterview data from a panel study, one can think of it as a set of related regressions or as a problem of studying individual change. And if there are more than two observations, one may want to treat each pair of periods as a set of change observations or take a longer period of change to reduce the number of potential changers who did not reveal it. Measuring change independently of level is of course necessary because of scale effects, real regression to the mean, and spurious regression from errors in measurement, so we use the average level or a lagged dependent variable among the predictors.

Finally, when there are many categorical classifications, including the dependent "variable" if any, then Leo Goodman (1972 and 1973) has developed a chi-square analogue to the usual variance approach, allowing one to apply all possible tests to a complete hierarchical multiway cross classification.

The most important problem in all this array of statistical approaches, however, is that many of them are designed for, and others are used for, ransacking data in a search for one or a few candidate models, yet we often want some unbiased assessment of the significance, explanatory power, and expected effect of each explanatory variable. That requires preserving an independent subsample not used in the search process. *We now issue our data with four independent subsamples indicated.* (With complex samples, a random subsample is not independent because of the clustering, so only the sampler can designate independent part-samples.) How much is kept for the assessment depends on the relative importance of selecting a model versus estimating its parameters and its power.

IV. Conclusion

I hope all this discussion of methods will not discourage anyone from getting some of our data and working with them. A great deal can be learned with relatively simple methods, and most of the refinements affect estimated sampling errors far more than they affect estimated relationships.

One reason those of us on this program are eager to see more economists analyzing the available data is that we are concerned about the vast unmet needs for more and better survey data. We frankly hope that if more people start using the existing data, the needs and uses of additional data will be more apparent and the support for their collection more enthusiastic. Three examples: (1) a decade of wealth accumulation, wild changes in returns in different kinds of assets, inflation, and massive shifts of portfolios has gone by without any decent information about what has been motivating people or what they plan to do next with their savings. (2) We talk of urban development, public services, housing policy, and trickle-down, but there are no national studies on what people want, pay for, and get in housing, neighborhood, and community. Without some such bench-mark measurements, assessing who benefits from new construction seems bound to be treacherous. (3) Earlier studies seemed to show a massive trend toward earlier retirement, and the young planning still earlier retirement, but no one has remeasured the plans nor found what the current inflation is doing to them. I could go on, but you can supply your own agenda. The real problem is that behavioral economics, to which we really must turn to tighten our models, must mostly make use of the natural experiments of history. This means collecting data with a calculating look at what has been going on and what we expect to happen next.

REFERENCES

F. M. Andrews and R. C. Messenger, *Multivariate Nominal Scale Analysis: A Report on a New Analysis Technique and a Computer Program,* Ann Arbor 1973.

———, J. Morgan, J. A. Sonquist, and L. Klem, *Multiple Classification Analysis,* rev. ed., Ann Arbor 1974.

R. E. Barfield, *Early Retirement, The Decision and the Experience and A Second Look,* Ann Arbor 1970.

——— and J. Morgan, *Early Retirement,* Ann Arbor 1969.

R. Barlow, H. Brazer, and J. Morgan, *Economic Behavior of the Affluent,* Washington 1966.

A. Conard et al., *Automobile Accident Costs and Payments,* Ann Arbor 1964.

J. Dickinson, "Labor Supply of Family Members," in *Five Thousand American Families—Patterns of Economic Progress,* Ann Arbor 1974, *1,* 190.

M. W. Gillo and M. W. Shelly, "Predictive Modeling of Multivariable and Multivariate Data," *J. Amer. Statist. Ass.,* Sept. 1974, *69,* 646–53.

T. K. Glennan, Jr., "Using Experiments for Social Research and Planning," *Monthly Lab. Rev.,* Feb. 1972, 16–19.

A. S. Goldberger and O. D. Duncan, eds., *Structural Equation Models in the Social Sciences,* New York and London 1973.

——— and R. Hauser, "The Treatment of Unobservable Variables in Path Analysis," *Sociological Methodology 1971,* San Francisco 1971.

L. Goodman, "A General Model for the Analysis of Surveys," *Amer. J. of Soc.,* 1972, *77,* 1035–86.

———, "Guided and Unguided Methods for the Selection of Models for a Set of T Multidimensional Contingency Tables," *J. Amer. Statist. Ass.,* Mar. 1973, *68,* 165–75.

G. Hendricks and K. C. Youmans, *Consumer Durables and Installment Debt: A Study of American Households,* Ann Arbor 1973.

J. B. Lansing and G. Hendricks, *Automobile Ownership and Residential Density,* Ann Arbor 1967.

———, C. W. Clifton, and J. Morgan, *New Homes and Poor People: A Study of Chains of Moves,* Ann Arbor 1969.

——— and J. Morgan, *Economic Survey Methods,* Ann Arbor 1971.

———, R. Marans, and R. Zehner, *Planned Residential Environments,* Ann Arbor 1970.

J. Morgan, "How Useful is the Cross-Section Survey?" *Monthly Lab. Rev.,* Feb. 1972, 3–10.

———, "Gasoline Price Inflation," *Econ. Outlook USA,* Spring 1974, *1,* 7–8.

——— and R. Messenger, *THAID, A Sequential Analysis Program for the Analysis of Nominal Scale Dependent Variables,* Ann Arbor 1973.

——— et al., *Five Thousand American Families—Patterns of Economic Progress,* Ann Arbor 1974.

———, I. Sirageldin, and N. Baerwaldt, *Productive Americans,* Ann Arbor 1966.

———, M. David, W. Cohen, and H. Brazer, *Income and Welfare in the United States,* New York 1962.

E. Mueller, *Technological Advance in an Expanding Economy: Its Impact on a Cross Section of the Labor Force,* Ann Arbor 1969.

M. Nerlove and S. J. Press, *Univariate and Multivariate Log-Liner and Logistic Models,* Rand Report R-1306-EDA/NIH, Santa Monica, Dec. 1973.

H. S. Parnes, "Longitudinal Surveys: Prospects and Problems," *Monthly Lab. Rev.,* Feb. 1972, 11–15.

J. A. Sonquist, E. L. Baker, and J. Morgan, *Searching for Structure,* rev. ed., Ann Arbor, 1974.

A. Zellner, "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *J. Amer. Statist. Ass.,* 1962, *57,* 348–68.

Potential Applications of Census Bureau Economic Series in Microdata Analysis

By SHIRLEY KALLEK*

Even today most economic research related to the enterprise or firm is limited to the utilization of aggregated data series. In fact there are few economic models in use today which do not depend upon macro rather than microdata sets. In this context, macrodata are defined as aggregated statistical series and microdata as information about an individual firm. Even though the theory of value is dependent upon decisions made by the individual enterprise, most econometric models deal with macro relationships. The relationships in any macro series are normally complex in nature since highly aggregated series reflect both actions of many different individual units and the shift in distribution of those units.

The basic questions that arise seem to be: Will analysis of microdata sets prove more useful in explaining economic phenomena than aggregated series? Will the integration of microdata sets with macro series enhance economic research and shed more light on existing economic theory? Although the consensus from most available literature indicates that the answer is in the affirmative, relatively little progress has been made in extending the use of microdata analysis. The Census Bureau, probably more than any statistical organization, has the most complete set of information about the characteristics of enterprises and their establishments—data which come from quinquennial census programs, from annual reports, and on a more limited basis from monthly and quarterly surveys. Moreover these data are available over long periods of time. The usefulness of this large body of data could be enhanced greatly through analysis at the micro level, i.e., through studying the behavior of the individual firms which comprise the statistical series. I would like to talk about some of our past efforts in exploiting this information base, why they were not satisfactory, and what our current plans are in this area.

In general, at the Census Bureau the use of microdata as a tool for fundamental research about the enterprise or economic reporting unit has lagged behind the progress made in the demographic area. The primary reason I believe for not tapping the vast store of microdata available in the economic sector has been its general unavailability to most researchers because of confidentiality restrictions. Reports filed by business concerns are strictly confidential by law and can be used only by employees of the Census Bureau for statistical purposes. We have not yet found any satisfactory means of issuing a public-use tape of individual establishments or firms which would eliminate the identifying factors of the individual firm and still permit the information to be useful. The highly skewed distribution of business establishments is not similar to the data in the demographic areas. Also utilizing a time series file of establishment reports presents new problems in terms of more subtle disclosure possibilities than the Bureau has ordinarily had to face in publishing data for single time periods.

---

* Associate Director for Economic Fields, U.S. Bureau of the Census. I would like to thank Thomas Mesenbourg and William Menth of the Census Bureau staff for providing assistance in the preparation of the special tabulation.

This, in effect, means that any research effort would have to be performed here at the Census Bureau or under contract by Census Bureau employees, and that the resulting analysis would be released only after it had been reviewed for possible disclosure. We have used both methods in the past for the small number of microdata studies we were involved in. Some possible alternatives will be discussed later in this paper.

Even if the problems associated with disclosure were resolved, there would remain other significant cost and data problems. Most researchers attempting to use microdata files have had frustrating experiences. This is still an area where much more knowledge must be obtained before this type of analysis can be optimized. The cost and complexity of handling these massive data files are enormous for any one individual or even for most groups. Then there are the numerous problems associated with utilizing existing microdata sets which are obtained as a byproduct of survey collection. Many of the issues and problems associated with the utilization of existing microdata sets for economic research have been discussed by F. T. Juster, G. H. Orcutt, and H. W. Watts, as well as Nancy and Richard Ruggles. It is worthwhile to discuss some of these problems since they are typical of the difficulties encountered when attempting to use any set of data tapes for purposes other than were originally intended. If they can be solved, significant progress can be achieved.

The Census Bureau has, naturally, placed its emphasis on the collection of data for the compilation and publication of statistical surveys and truthfully has not given much attention to the use of these data as inputs for microdata analysis. Thus the entire systems design, review, and processing phases for any survey are generally not compatible with the analytical needs. For example, in most surveys designed by the Bureau, the data base file covers only the two periods, current and previous, which are needed for the review and publication of the series. The volatility of the economic reporting units (from mergers, closings, etc.) results in numerous changes in the identification numbers of the reporting units. For purposes of the survey review, comparability of the reporting units and a common identification number need be maintained only for the two periods; for economic research purposes, this comparability must be extended over the period of the entire file.

Questions also arise as to the comparability of reporting units. Should changes reflected in the reporting unit as the result of industry code changes, ownership changes, etc. result in treating the reporting unit as a discrete entity in each survey year, or should such results be linked? If they are linked, the disclosure problem is increased enormously. In some instances sampling rules for surveys will treat a change in ownership as a "death" and in others as a "merger" of the acquiring firm. Thus there may be inconsistencies from survey to survey depending upon the sample design and requirements of the survey. Moreover the changing identification number due to mergers, changes in corporate structure, etc. results in difficulties in matching records from one time period to another.

There are also omissions or data errors in the records of individual establishments. Some have not been detected and never will be. The review of data for publication totals can be quite different from that which is needed for manipulation of the information at the micro level. Our review process does not "clean" the data beyond the point needed for tabulation analysis. It is the hope of every statistical agency that any remaining errors are minor. More important are the errors which are detected during the survey processing, but which, primarily because of budgetary reasons, are not carried back to the computer tapes although the corrections are reflected in the published data.

Many Census Bureau surveys are based on sample panels which are updated at periodic intervals. A continuous time series for a long period of time will therefore of necessity be limited to the larger establishments. The practice of completely changing the noncertainty portion of the sample to minimize reporting burden aggravates the problem of comparability for the researcher. These are just some of the examples which can affect any examination of the microdata. If research utilizing these data is to be encouraged, accommodation to these conditions must be made in the design of the operating system.

In the early 1960's, the Bureau attempted a major longitudinal study of economic establishments, but the results were less than encouraging, primarily due to inexperience in resolving the problems just discussed. Several years of work went into creating a file of statistical histories for individual manufacturing plants, primarily using data reported in the Annual Survey of Manufactures.

Even though these basic files were difficult to use and were deficient in many ways, several studies utilizing these data were undertaken. One was the Ruggles project, "The Behavior and the Characteristics of Manufacturing Establishments in the U.S. 1954–1959," undertaken under the joint sponsorship of the Social Science Research Council and the Bureau of the Census, which related to the price-cost behavior and characteristics of manufacturing establishments. Another outgrowth of the establishment time-series project was a pilot study on measuring capacity through the peak performance potential approach. In addition, the Bureau of Labor Statistics conducted a methodological study to explore whether plants with high productivity levels could serve as a guide to projecting future productivity increases of an industry.

The major lesson learned from the 1960's time-series effort is that it is almost impossible to reconstruct an historical file of individual establishment records. The tracing of identification numbers for individual establishments and the *ex post facto* correction of data fields pose processing problems which are almost impossible to overcome.

If the Bureau is to provide the ability for microdata analysis, the concept must be built into its operating system, and research efforts must be undertaken here at the Bureau which will permit optimum manipulation of the data. Alternatives to the issuance of public-use tapes must also be explored, since our research over the past twenty years has not resulted in any significant progress in resolving the disclosure problem. We have established an economic research unit whose responsibility will be to prepare microdata analysis for others on a reimbursable basis. This hopefully will resolve some of the operational problems of the past where we tried to have these studies prepared by the same operating staff responsible for the survey management.

This unit, under Bureau sponsorship, has undertaken a research effort which will attempt to match information from our demographic files with our economic files. In particular, the Bureau has at its disposal information about individual workers. We are attempting to see if it is possible to match the individual characteristics of employees with the operating characteristics of the establishments in which they are employed, interrelating such variables as age, race, and income of employee with size of the firm, its industrial classification, its geographic area, capitalization, etc. Work is just beginning in this area and the success of the project is still not known. If fruitful, it can open up an entire new area which has never been tapped.

We also plan to explore the development of a wide range of variance-covariance matrixes which could be made available to outside researchers. Initial review has indicated that by using a sufficiently large number of observations and by withholding some of the constants it may be possible to release such matrixes without problems of disclosure.

Another new approach, in effect, will integrate the needs of microdata analysis into the data system. We think that this is now possible with the introduction of an industrial directory system here at the Bureau. The Industrial Directory which has been under development for the past several years is now becoming operational. The Directory, which is confined to Bureau use, is basically a computerized file of all companies and their constituent establishments. Fundamental to this system is the inclusion of all changes in ownership, linkage of identification numbers, changes in industry classification, etc. for all multi-establishment firms and for all firms reporting in the Bureau's current surveys. This file, while too large an entity to manipulate, will be the basis for maintaining the Master Sample. The Master Sample was designed originally to control the quality of the internal processing of the 1972 Economic Censuses program files and to identify reporting units which could be subjected to intensive quality checks both before and after the full-scale processing. While the main purpose was to prevent systematic errors in processing, the Master Sample is large enough to measure the various attributes of the universe with adequate reliability. The Master Sample can be a powerful tool for the analysis of economic microdata at reasonable cost.

The sample itself contains about 200,000 employer establishment records and includes all economic divisions. It basically provides for a 2 percent sample of all establishments with less than 250 employees and a certainty stratum of all employer records with 250 or more employees. For example, approximately 30,000 manufacturing plants, 10,000 wholesale establishments, and about 50,000 retail stores will be identified. It started with the 1972 report year and will be maintained on an annual basis. The noncertainty segment may not, however, be large enough for the manufacturing sector and a subsample from the Annual Survey of Manufactures panel may have to be included in the Master Sample. As a tool for analyzing economic microdata, the Master Sample lends itself to tracing individual cases over time at nominal costs. Changes in industrial activity, growth patterns, and geographic mobility can be followed by matching manageable record files from year to year and census to census. Moreover, correction of the data in the basic computer files is being built into our regular processing system.

In view of the considerable interest in this area, the Bureau is again giving considerable attention to ways of releasing new types of analytically useful data that can be derived from examining the interrelationships among the different variables of microdata sets. One of the possibilities is the publication by the Bureau of various analytical ratios and studies which would permit a deeper examination of the underlying data. There are many advantages to be gained in examining more detailed sets of the same group of aggregated data. First, underlying movements of economic phenomena may be masked by aggregative totals. Movements of larger plants can react quite differently from small ones. In fact, in many cases that is exactly what we want to find out. To show the value of such disaggregation, the establishments in the data files for the 1970–71 Annual Surveys of Manufactures were matched and analyzed to attempt to determine if there is a

TABLE 1—QUINTILE DISTRIBUTION OF ESTABLISHMENTS BY RATE OF CHANGE IN PRODUCTIVITY (1970–71) AND EMPLOYEE SIZE FOR MANUFACTURING PLANTS

Quintile	Mean Rate of Productivity Change (1970–71)	Percentage of Establishments in Quintile (By Employee Size Class of Plant)		
		1–49 Employees	50–249 Employees	250 or More Employees
1	.96	25.4	19.7	16.8
2	1.05	19.9	21.4	20.9
3	1.09	19.7	19.7	23.1
4	1.12	16.1	20.7	22.3
5	1.21	18.9	18.5	16.9
Total	1.08	100.0	100.0	100.0

Source: Unpublished tabulations from 1970–71 Annual Survey of Manufactures, U.S. Bureau of the Census.

direct relationship between the size of an establishment and its rate of change in productivity. For example, the dollar volume of adjusted value added (*AVA*) was used as the measure of output and the productivity for each establishment was computed by dividing *AVA* by the total manhours worked for both years and the rate of change from 1970 to 1971 was derived.

In interpreting the rates of change a word of caution must be given. The productivity measures used in this example are rather crude and are based on dollar volume of output, rather than physical output. Unlike physical output, value of output is affected not only by difference in input costs, but also by differential demand factors for different industries and for different establishments within the same industry. In addition, total manhours worked are affected by capital input.

The establishments within each of the 420 manufacturing industries were ranked by rate of change in *AVA* per manhour, placed in approximated quintiles in ascending order according to their rate of change from 1970 to 1971, and sorted by three employee size classes. While the overall mean rate of change was 1.08, the mean rate of change varied from .96 for the first quintile to 1.20 for the fifth quintile.

As seen in Table 1, there are significantly different patterns for each of the three size groups. The largest employee size category appears to be almost an inverted U-shaped distribution; a larger number than would be expected fall in the second, third, and fourth quintiles. For the smallest size category, it is evident that a relatively higher number of establishments fall into the first quintile. For the middle range, i.e., 50–249 employee size category, the differences among quintiles are less pronounced. It appears that the larger establishments are much more stable and are less likely to experience extreme increases or decreases in rates of productivity change. The smaller establishments, i.e., under 50 employees, are much more likely to experience very small or even declining rates of productivity. Further investigation would be necessary to see what factors influence these changes.

It would be interesting to correlate the rates of change with gross assets or changes in gross assets of the establishment to examine productivity and investment behavior. Or, if data for additional years were linked to the 1970–71 file, would establishments increasing in size have different productivity rates compared to those that remain fairly stable in size?

Needless to say, this is just a fairly simple example of the types of research which can be undertaken when using microdata. Once the basic file is extended to include a longer time period, the possibilities for research at this detailed level can become even more meaningful. Hypotheses can be tested more fully and changes over time can be examined for any number of variables, and the behavior of individual establishments at different phases of the business cycle can be studied. The possibilities are almost limitless for there is a body of data waiting to be analyzed.

REFERENCES

M. R. Conklin, "Time Series for Individual Plants from the Annual Survey of Manufactures and Related Data," *Amer. Statist. Ass. Proc. of the Bus. and Econ. Statist. Sec.*, 1964, 404–10.

W. K. Jordan, "The Measurement of Performance Potential in Manufacturing Establishments," *U.S. Bur. of the Census*, work. pap. no. 18, 1965.

F. T. Juster, "Microdata, Economic Research, and the Production of Economic Knowledge," *Amer. Econ. Rev. Proc.*, May 1970, *60*, 138–48.

———, "Microdata Requirements and Public Policy Designs," *Annals of Econ. and Soc. Measurement*, Jan. 1972, *1*, 7–16.

G. H. Orcutt, "Basic Data for Policy and Public Decisions: Technical Aspects—Data, Research and Government," *Amer. Econ. Rev. Proc.*, May 1970, *60*, 132–37.

N. Ruggles and R. Ruggles, "The Macro Accounts and Microdata Sets," *Amer. Stat. Ass. 1970 Proc. of the Bus. and Econ. Statist. Sec.*, 208–13.

———, "Price-Cost Behavior of Manufacturing Establishments," unpublished paper, Mar. 1965.

H. W. Watts, "Microdata: Lessons from the SEO and the Graduated Work Incentive Experiment," *Annals of Econ. and Soc. Measurement*, Apr. 1972, *1*, 183–91.

The Experimental Panel Data Resources at the Poverty Institute Data Center

By KATHERINE DICKINSON AND HAROLD W. WATTS*

The Institute for Research on Poverty has been involved in two negative income tax experiments sponsored by the Office of Economic Opportunity: The New Jersey Income Maintenance Experiment in New Jersey and Pennsylvania (conducted jointly with Mathematica, Inc.); and the Rural Income Maintenance Experiment in North Carolina and Iowa. The information collected in these experiments form two very large panel data bases. The Data Center was established with support from the office of Health, Education and Welfare to facilitate and encourage research on these data by providing consultation, documentation, and data files to potential users. The Data Center is also planning to provide a combined data base by merging the comparable questions used in both the urban and rural experiments.

The Data Center was created in the belief that full scientific and social benefits from the large expenditure of effort and resources which the experiments represent can only be realized from extensive and critical analysis from many points of view. While there is no way of assuring that this will happen, there are substantial obstacles in the way of using large and complex data files that can be removed. In addition it is possible to use every available opportunity to encourage, inspire, or provoke research scholars to engage in the challenging task of exploiting a new and, in part, unique body of evidence. In this presentation we aim at creating an interest in using the experimental data, and indicating how the Data Center can ease the always difficult job of getting on working terms with a new data file.

I. The Experiments Briefly Described

The New Jersey Income Maintenance Experiment was the first large-scale social experiment conducted in this country. It was a carefully controlled field test on the effects of eight negative income tax plans. Such a plan can be defined by the guarantee level (the level of payment the family receives if it has no other income) and the tax rate (the rate at which payments are reduced for each dollar of other income). Various combinations of four guarantee levels and three tax rates were tested. Table 1A shows the number of families on each plan for the urban experiment. Experimental families received biweekly checks reflecting their plan and their three-month average income; control families received nominal amounts for participating in the interviews.

The sample included 1,357 families who lived in one of four Northeast metropolitan areas (Trenton, Jersey City, the Paterson-Passaic *SMSA* in New Jersey, and Scranton, Pennsylvania). Each family initially contained an able-bodied man aged 18 to 58 (in most cases the head of the household) and had income of less than 150 percent of the official poverty line. This experiment concentrated on the effects on men, since few other transfer programs extended eligibility to them and thus little was known about their work-effort response to transfers. The experiment lasted

---

* Director, Poverty Institute Data Center, and professor of economics, University of Wisconsin, Madison, respectively.

TABLE 1A—ALLOCATION OF FAMILIES TO EXPERIMENTAL PLANS IN THE NEW JERSEY INCOME MAINTENANCE EXPERIMENT

Plan (Percent Guarantee and Tax Rate)	Income Stratum Low	Medium	High	Total
50–30	5	31	12	48
50–50	29	37	5	71
75–30	35	14	55	104
75–50	5	67	46	118
75–70	18	66	0	84
100–50	22	39	20	81
100–70	11	26	43	80
125–50	50	8	80	138
Total Experimental	175	288	261	724
Control	238	165	247	650
Total Sample	413	453	508	1,374[a]

[a] This number differs from the actual sample size of 1,357 because of sampling shortfall.

three years in each site, starting in the summer of 1968 in Trenton and ending in the fall of 1972 in Scranton.

The rural experiment was a similarly designed study which involved 809 families living in the rural South and Midwest (Duplin county in North Carolina and Pocahontas and Calhoun counties in Iowa.) There were 217 farm families included in the sample. This experiment studied eight plans, six of which were also used in the urban experiment. The allocation of families to various plans is shown in Table 1B.

The majority of the rural sample consisted of families with male heads between 18 and 58, but there were 108 nonaged female-headed families and 114 families with heads over 58. The income cutoff for the rural experiment was also 150 percent of the poverty level. The experiment was conducted from early 1970 to mid-1973.

This necessarily abbreviated description of the experiments can be extended by any or all of the items listed at the end of this

TABLE 1B—ALLOCATION OF FAMILIES TO EXPERIMENTAL PLANS IN THE RURAL INCOME MAINTENANCE EXPERIMENT

Plan (Percent Guarantee and Tax Rate)	Income Stratum Low	Medium	High	Total
50–50	9	42	0	51
75–30	11	0	73	84
75–50	5	39	55	99
75–70	8	34	0	42
100–30	1	1	3	5
100–50	6	0	75	81
125–50	1	0	6	7
125–70	1	3	1	5
Total Experimental	42	119	213	375
Control	38	140	257	435
Total Sample	80	259	470	809

paper. References dated 1973 or later include experimental findings as well as some descriptive accounts. The Users Manuals for the Data Files, described below, also contain complete descriptions of the experimental design and procedures.

II. Nature of the Data Collected

The families in both experiments were given a baseline interview, twelve quarterly interviews during the operation of the experiment, and a follow-up interview three months after the last payment was made. Approximately 3,000 variables were measured, and the information collected covers many aspects of behavior, attitude, and circumstance.

The primary goal of the experiments was to measure the work-effort response to both an increase in income and a decrease in the marginal wage rate. To monitor changes in labor supplied, each quarterly interview contained a core set of questions on earnings, hours of work, and other labor force behavior (see Table 2). These core questions were asked of up to five adults in each family in the urban experiment and of all adults in the rural families. The questions common to both studies for all quarters concern the previous week's hours, earnings, and occupation. The rural study also concentrated on farm income and expenses each quarter.

The noncore section of the interviews contained questions on economic, social, and psychological aspects of the families' lives which either could be changed by the experiment or could influence the effect of a negative income tax. These questions were rotated among the quarters and were generally asked two or three times. Tables 3 and 4 contain lists of the major categories of the information obtained.

Three general characteristics of the data deserve special mention here. First of all, they are panel data using the household-family as the primary observational unit, but allowing substantially complete derived panels of individuals to be assembled as well. As with any panel data file, there

TABLE 2—CORE VARIABLES COLLECTED EVERY QUARTER IN THE EXPERIMENTS

New Jersey Experiment	Rural Experiment
Family composition	Same
Labor-force experience last week Hours—main job, overtime, second job, hours lost Earnings Occupation and industry	Same
Labor-force experience last month	Labor-force experience last thirteen weeks
Sources of unearned income	Same
Housing costs	Same
—	Migration
—	Farm and business income and expenses
—	Unpaid hours in farm or business for women
Travel to work	—
Child care and support	—

Table 3—Other Categories of Data Collected on the New Jersey Income Maintenance Experiment

Attitudes
—anomie, control of future, happiness, ladder of life, worry, self-esteem, social desirability bias, time horizon, future events, political liberalism, awareness, and integration

Education and Ability
—educational attainment of children and parent's aspirations, education of head and wife, education of parents and siblings, Ammons Quick Test for head

Family Information
—composition, social and demographic characteristics, marital history, religious background, family integration, homework and children's chores, child care, child support, family planning

Financial Information
—expenditure patterns of food and clothing, installment buying, debit, savings, durable goods, and insurance

Housing and Property Ownership
—housing status and costs, housing conditions, home improvements

Job History
—lifetime job history of head, job history of wife since current marriage, occupational experience of parents and siblings

Job Satisfaction
—attitudes toward job mobility, job expectation, job satisfaction and alienation

Job Training Programs
—participation of head and wife

Leisure Activities
—time-use study, family activities, sports, hobbies and vacations

Media Use

Medical and Dental Experience
—attitudes, psychosomatic scale, current care, long-term disabilities, hospitalization, dental care

Group Involvement
—organizational membership, social integration, neighborhood ethnicity

Residential Mobility

Welfare Experience
—history and during experiment

Table 4—Other Categories of Data Collected on the Rural Income Maintenance Experiment

Attitudes
—life satisfaction, self-esteem, self-comparison, ladder of life, personal efficacy, future orientation, family loyalty, anomie, social desirability bias, time preference

Political
—efficacy of and attitudes toward central government, political behavior

Education and Ability
—expectations for children, educational history, children's performance, Ammons Quick Test

Family Information
—family planning, family relations, division of labor, marital history, social and demographic characteristics, child care

Annual Finances
—income, net worth, investments

Farm
—government programs, crops, yields, tenure, livestock

Other Financial Information
—credit and debt, insurance, consumer goods expenditures

Food and Nutrition
—24-hour intake recall, food expenditures and budgets, food preparation, food subsidies, attitudes about food and cooking, food during pregnancy

Housing
—quality, costs, and home repairs

Job History and Attitudes
—job history, attitudes toward job mobility, job satisfaction, attitude toward working wives, work-related and job-search expenses, frequency of pay, fringe benefits, travel to work, work orientation, children's jobs, desired work hours, discrimination

Health
—use of health facilities, current state of health

Activities
—social integration, organizational involvement, leisure activities and vacations, religious activities

Public Programs
—attitudes, knowledge, and use of welfare, social services and agricultural programs

Teenage Supplement
—attitudes toward school, aspirations for school and job, social and family integration, self-esteem, psychosomatic scale, delinquency

are problems with interrupted and incomplete data series for some persons or households. But there is also the opportunity for longitudinal analysis based on intertemporal variation which can, as in any panel data, be compared with the interunit variation of the cross section.

A second and unique feature of the data lies in the frequency of interviewing. Almost all currently available panel surveys are annual or less frequent. The experimental panels were interviewed quarterly and consequently the data can be used for examining dynamic adjustment processes which require more closely spaced measurements.

Finally, the panel data records the behavior of a sample which was exogenously "treated" with various "doses" of income subsidy on a controlled and randomized basis. This is, of course, the principal unique feature of this data resource. This feature affords the possibility of examining the consequences of experimental variation and to contrast them with the effects of natural or spontaneous variation observed (across individuals or over time) in nonexperimental panels or in the control group. It is hoped that, as for experimentation in the physical sciences, the introduction of clearly exogenous variation will simplify the identification of causal relations.

III. Uses of the Data

The estimation of the labor supply response to the guarantee levels and tax rates was the major focus of the experimental design and of the initial analytic reports. These studies provide evidence on the work-effort effects for wage earners, the self-employed, wives and female heads, and other family members. Information was also collected that makes possible the study of experimental effects on wage rate. A variety of such effects might be expected. If a worker does reduce his hours of work, it may result from not working overtime or on a second job, both of which usually have a wage rate different from the main job. Information on these work hours is available every quarter. Having a floor on income also reduces the risks involved in changing jobs. Extensive job histories, both before and during the experiment, were collected for the head and wife and can be used to test this hypothesis. A guaranteed income would also make job training more attractive. Prior education and participation in training programs during the experiment were asked of the head and wife.

The farm families might be expected to engage in more risky enterprises as a result of the experiment or increase specialization in one product. The guarantee might also provide a floor on farm investments in such things as fertilizer which have a very high payoff.

If families do reduce the number of hours worked in the market, it is important to determine what the family does with the increased time. The parents may, for example, devote more time to their children, resulting in a more integrated family and in the children doing better in school. Both experiments contain measures of time-use, family integration, and children's educational attainment; the rural data even contain information gathered from the schools on the children's performance.

Unlike food stamps or housing allowances, a negative income tax places no constraints on how income is spent. An important aspect of these studies is to determine how the payments were used and specifically whether these transfer plans achieved the goals of the in-kind transfer programs. Data on housing, food, and clothing expenditures were collected as well as purchases of furniture, appliances, and vehicles. These durable goods expenditures and the data on savings and debt can also provide insights into the validity of the permanent income hypothesis and evidence on the short-term nature of the experiment.

Health can affect the labor supply response and can itself be changed by the experiment. The latter response would be expected to include a substitution effect as well as an income effect, since time is an important cost of health care. A substantial amount of health care and disability information was collected on the husband, wife, and children, particularly in the urban experiment.

Social psychological attitudes also may interact with labor force behavior and be themselves changed through a negative income tax. It may be that only families poorly integrated into the major values of society would reduce work effort. Further, the increased income and reduced stress resulting from a guaranteed income might increase self-esteem, feelings of personal efficacy, and the quality of life of the recipients. The result might also be more integration into society. These hypotheses, similar to some used in the Survey Research Center's Panel Study of Income Dynamics, are testable with these data.

Other aspects of social behavior which could respond to the experiment involve the composition of the family itself. Increased stability might be expected because coverage is extended to intact families under these plans, in contrast to current transfer systems for which male-headed families are often ineligible. Further increased income and less variability might produce less family stress. Family planning decisions could be affected not only from increased income but also because benefits are tied to family size and because the cost of time inputs to children is reduced.

Many of the above hypotheses have been given a first examination in the final reports for these experiments and in other research. The bibliography below provides references to available materials and the Data Center endeavors to provide references to work completed or in progress on various subjects for those interested in pursuing a topic.

IV. Obtaining the Data

In general, the complete data sets from the experiments are not available. They are simply too large and too complex to be distributed on a routine basis. Instead, the Data Center can provide a tailor-made subset of the data, according to the user's specifications. For the urban experiment, the data tape used in most of the economic analysis for the final report is also available. The project to merge the two data bases is still in the planning stages but information should be available by mid-1975.

The Data Center has compiled User Manuals which document the experiments (including study design, sample selection, the rules of operation, and methodological issues) and the data (detailed descriptions of the variables collected and instructions for specifying extracts). These manuals can be obtained from the Data Center upon request. After consulting the manual the user can then order (without charge) code books for all the possibly relevant categories of data, which contain the actual questions asked, the quarters each is available, and the coding sets used. From these the desired variables can be specified. Charges for extracts are at marginal cost and generally are less than $300.

The address of the Data Center is as follows:

Data Center
Institute for Research on Poverty
University of Wisconsin
1180 Observatory Drive
Madison, Wisconsin 53706

General or specific inquiries are cordially invited and suggestions as to how we can be of greater assistance to research analysts are always welcome.

REFERENCES

D. L. Bawden, "Income Maintenance and the Rural Poor: An Experimental Approach," *Amer. J. of Agricultural Econ.*, August 1970, *52*, 438–41.

——— and William H. Harrar, "The Use of Experimentation in Policy Formulation and Evaluation," *Urban Affairs Quarterly*, June 1972, *7*, 419–30.

D. N. Kershaw, "Negative-Income-Tax Experiment," *Scientific American*, Oct. 1972, *227*, 14, 19–25.

H. W. Watts, "Graduated Work Incentives: An Experiment in Negative Taxation," *Amer. Econ. Rev. Proc.*, May 1969, *59*, 463–72.

Institute for Research on Poverty and Mathematica, Inc., *New Jersey Graduated Work Incentives Experiment: An Informal Compiled Report*, Madison 1973.

U. S. Dept. of Health, Education and Welfare, *Summary Report: New Jersey Graduated Work Incentive Experiment*, Washington 1973.

Journal of Human Resources, "Symposium Articles—The Graduated Work Incentive Experiment," Spring and Fall 1974, *9*, 155–278, 502–55.

THE BEHAVIOR OF FIRMS IN OLIGOPOLISTIC MARKETS: IN THEORY AND PRACTICE

Firm Decision-making Processes and Oligopoly Theory

By PAUL L. JOSKOW*

The typical industrial organization economist interested in examining the behavior of firms in market environments characterized by small numbers does not approach his task with any unified set of analytical tools which one could call the theory of oligopoly. Instead, he comes armed with a whole smorgasbord of formal models, *ad hoc* models, case-study information, stories, and vague notions concerning the impact of business psychology and sociology. The various theories, whether they be formal or informal, are based on a bewildering collection of a priori behavioral assumptions, formal mathematics, hard case-study information, casual empiricism, hand waving, and as much as possible of the traditional theory of the atomistic competitive firm.

For those of us interested in public policy analysis this state of affairs is troublesome. On the one hand, many of the interesting policy issues arise in markets dominated by a small number of large firms and therefore some viable alternative to the competitive model is desirable. On the other hand, existing formal models often do not provide a useful framework for policy analysis.

This essay proceeds by briefly reviewing the formal models which appear to make up the corpus of formal oligopoly theory. These models are then evaluated in terms of their ability to generate testable hypotheses that can differentiate one model from the next and oligopoly behavior from competitive market behavior, as well as the usefulness of these models in the analysis of particular markets and particular public policies. I conclude that there is little qualitative difference in the implications associated with most of the models, but, more importantly, that the formal models are not really utilized by serious students of actual markets and public policies. Rather, the important characteristic of much of this applied work is the use of informal models, "stories," the consideration of particular decision-making processes, situation-specific considerations of uncertainty, information costs, other transactions costs, and various institutional constraints. Since it appears that the important characteristics of oligopoly behavior are not captured by conventional models, some suggestions for further research which deals more explicitly with the behavior associated with the "oral tradition" of industrial organization are given; specifically, more careful analysis of actual firm decision-making processes and the phenomena which determine them is called for.

I. Formal Models

The formal modeling of duopoly and oligopoly has a long tradition. Virtually all of the modeling efforts (a) assume that firms seek to maximize profits and (b) make some behavioral assumption concerning the ways in which a particular firm relates to the others (actually or potentially) in the market. What makes oligopoly theory different from competitive market theory is the recognition by firms that outcomes of their decisions are affected by the behavior of other firms. However, the picture of the firm painted in most formal oligopoly models is really not much richer than that of the "black box" competitive firm. The major difference is the replacement of the obscure Walrasian auctioneer of competitive market models with specific price adjustment rules for individual firms. Exactly how different the resulting market behavior is depends upon the particular interrelationships that are assumed to characterize the environment in which a particular firm attempts to maximize profits.

---

* Massachusetts Institute of Technology.

Within the formal modeling tradition two basic trusts have emerged. While both make the assumption that firms attempt to maximize profits, they are characterized by different types of behavioral assumptions. The first set of models is characterized by behavioral assumptions concerning the ways in which one firm perceives or responds to the actions of others already existing in the market (W. Fellner). The typical model specifies a set of behavioral assumptions, derives the profit-maximizing conditions given these assumptions, and examines the nature of the resulting equilibrium in a market with two or more firms. The work of Augustin Cournot, J. Bertrand, F. Y. Edgeworth, and H. von Stackelberg exemplifies this tradition of formal modeling. The kinky demand curve model attributed to Robert Hall and Charles Hitch and to Paul Sweezy is often regarded as an approach in itself; however, the basic methodology is very much in the same tradition as the others.

The second formal modeling approach essentially ignores the interdependence of firms in existence and concentrates instead on the behavior of a firm faced with potential entrants into the industry. Such models either treat the existing market structure as pure monopoly or assume implicitly that existing firms have been able to collude and can coordinate their pricing or output behavior. These "limit-pricing" models make assumptions about the ways in which potential entrants perceive the response to entry by existing firms (they might be assumed to maintain their output but lower their prices in the face of entry) and determinants of the rate of entry, and recognize explicitly that entry is neither generally completely blockaded nor completely "easy." In the most simple models, the existing firm chooses some limit price, generally above the competitive price, but below the short-run profit-maximizing price, at which potential entrants would find it unattractive to enter the industry. The work of Roy Harrod, H. R. Edwards, Joe Bain, Paolo Sylos-Labini, F. Modigliani, and J. Bhagwati is characteristic of this literature.

Recent extensions of the limit-pricing idea have introduced concepts of entry uncertainty, the recognition that entry takes time and will occur at different rates depending on expected profits, and the opportunity to choose a trajectory of prices over time rather than a particular preentry limit price. The works of D. Gaskins, M. Kamien and N. Schwartz, and D. Baron are examples of recent extensions of limit-pricing notions recognizing uncertainty and dynamics more explicitly. While there have been a few attempts to bridge the gap between the two thrusts of formal oligopoly modeling (F. Fisher and C. Frank, Jr.), extensive efforts in this direction have not been made.

A natural question to ask is what empirically testable hypotheses can be derived from these models. In particular, can they lead to hypotheses that can be differentiated from behavior implied by models of perfect competition, and can they be differentiated one from the other?

Virtually all of the models which attempt to recognize the interdependence of firms existing in the market lead to the conclusion that prices will lie somewhere between the competitive price and the short-run monopoly price in equilibrium, and that as the number of firms in the market increases, the equilibrium oligopoly price will approach the competitive price. The kinky demand curve model implies that prices in oligopolistic markets will be insensitive to moderate swings in the costs of production. While many problems exist with the kinky demand curve model as formulated (F. Scherer and G. Stigler) and while a number of researchers have purportedly disproven the validity of the empirical implications (Stigler, W. Simon, and W. Primeaux Jr. and M. Bomball), both the essence of the model and its empirical implications continue to be taken seriously by many economists. Given a continued belief in the "stylized fact" associating price rigidity with concentration, supporting empirical work (Stigler and J. Kindahl, p. 91), *and* general dissatisfaction with the kinky demand model itself, combined with the inability of the other traditional models to generate this empirical result, we must wonder why more effort hasn't been directed toward constructing a more satisfactory model. Except for the kinky demand curve model, however, the interdependence models are not particularly rich in empirical conclusions especially with regard to short-run dynamics and the characteristics of the path toward a new equilibrium.

The limit-pricing models suggest that the limit price will lie somewhere between the short-run monopoly price and the competitive price. The difference between the long-run equilibrium limit price and the competitive price varies directly with the scale at which minimum average cost is achieved, other barriers to entry, the discount rate, and the rate of growth of demand (Gaskins). It varies inversely with the elasticity of demand, the extent of the market, the probability and rate of entry (*ceteris paribus*), and the level of risk aversion. As a result we would expect to observe higher profits in markets with substantial scale economies and technological change, entry of new firms over time (as long as there is uncertainty about entry and entry is not instantaneous or "all or nothing") and a declining (or perhaps eventually constant) market share for the dominant firm.

All of these predictions are consistent with available empirical evidence. There seems to be a positive relationship between concentration and profitability (Bain, Michael Mann, R. Sherman and R. Tollison), as well as a positive relationship between profitability and barriers to entry (Bain and Mann). There is entry over time into many markets which historically had one large dominant firm (steel and tin cans are standard examples, but aluminum and shoe machinery are counterexamples) and the share of dominant firms has often tended to decline at least during some part of the history of the industry (some examples and counterexamples). In addition, we do observe price leadership and intermittent warfare in some industrial markets (the cigarette market is an example).

The limit-pricing models have been advertised as a great step forward, but the vast array of general empirical results that we have makes it extremely difficult to differentiate among these models as well as between limit-pricing models and interdependence models. Firms could be the most naive, obeying the Cournot assumptions and ignoring potential entrants, and we would still observe entry, declining profits and prices, and declining market shares over time. Whether the existing firms are rational limit pricers or not, excess profits will eventually attract entry, and the degree of entry will depend on how easy it is to enter. I suspect that most of the *qualitative* results of the limit-pricing theories would go through with a simple Cournot model plus entry. As far as I am concerned the limit-pricing models are more an addendum to the theory of monopoly, recognizing the possibility of entry and the optimality of long-run rather than short-run profit maximization, than a major insight into the economics of behavior in oligopolistic markets.

II. Formal Models and Empirical Analysis

In a sense the ultimate test of the utility of the various models is whether they prove useful to people involved in analyzing problems involving actual markets or groups of markets. I suggest that not only aren't they particularly useful but also that they really aren't even used.

Let's first consider empirical work that has been done to assess general relationships between market structure and behavior. Virtually none of it appeals to particular formal models of oligopoly markets. It recognizes that prices and profits should be positively correlated with such things as concentration, barriers to entry, product heterogeneity, etc., but the empirical specifications appeal more to a commonsense understanding of the way markets work in general rather than to particular models. Variables reflecting the behavioral assumptions specified in some models and found in certain markets, such as price leadership and price rigidity, are rarely used as explanatory variables. The oligopoly models have little or nothing to say about important issues, such as research and development efforts, technological change, wage and price adjustment rates, and technical efficiency, and how such behavior might differ between and among industries with competitive and oligopolistic market structures.

When we move from general empirical work designed to elucidate basic propositions about the relationship between structure and behavior to studies of particular markets, any facade that the investigation is based on some formal oligopoly theory or set of theories is dropped altogether. Serious investigators get right down and attempt to understand the actual interactions among firms, the ways in which decisions on prices, outputs, research and development activity, and entry are made, and the major decision variables used to signal to, coordinate with, and punish rivals who don't fall into line for the particular industries under study (C. Kaysen, J. Dirlam and A. Kahn, A. Chandler, R. Nelson (1962), and many others). Plausible and interesting stories about pricing, output, entry, and research and development decisions abound as a result of these studies.

Somehow one gets the distinct feeling that the important messages are being carried by the informal theories, stories, and behavioral observations, and that the formal models are trotted out *ex post* to demonstrate that some kind of formal apparatus can explain or incorporate some of what is actually being observed. The formal models often have great mathematical elegance and are fun to play with and teach, but at least from this author's vantage point their existence has had a negligible effect on applied research and policy analysis in the industrial organization area. General policy prescriptions and uncertainties regarding concentration, price rigidities, barriers to entry, market definition, tacit coordination, etc., seem to exist independently of the formal modeling tradition.

III. A Behavioral Approach

The major problem with the conventional models of oligopoly arises from an approach which takes as given the neoclassical black box profit-maximizing firm embedded in a competitive market and then tacks on simple behavioral rules to turn it into an oligopoly model. The interesting stories which characterize much of the research on particular markets and responses to particular policy changes seem to speak much more to the ways in which actual firms make actual decisions. This oral tradition of industrial organization digs into the actual organizational structure of firms, decision-making processes within firms, information flows, environmental perceptions, business psychology, and institutional constraints and recognizes that these considerations are what is important. Formal oligopoly theory seems to ignore what is actually going on.

Consider pricing and entry behavior. One of the things that is supposed to differentiate oligopolistic markets from competitive markets is that firms are price-makers rather than price-takers. A number of case studies and some more general empirical work has suggested that many firms use relatively simple rules of thumb for setting prices—markup rules, target normal rates of return, etc. (see Scherer, p. 173, and A. Kaplan, Dirlam and R. Lanzillotti). These pricing rules appear to be important determinants of short-run pricing behavior in oligopolistic markets during periods of rapid changes in demand, technology, and input prices. There is a need to understand the price adjustment process better and to make use of models of price formation to predict pricing behavior in oligopoly markets. What is the nature of these rules of thumb? Why might they be different for different products, different market structures, and industries of different vantages (M. Piore)? How important are internal signals such as inventories and input price changes relative to external signals like the prices and market shares of rival firms (especially in markets with differentiated products)? Do firms even think about *potential* entrants or do the realities of bounded rationality make the processing of internal information and the monitoring of a few rivals all that is really possible? It is apparent that pricing rules of thumb and more complicated price decision-making processes are in fact the reality in which most firms operate.

Entry both real and potential plays an important role in virtually all approaches to the theory of the firm and the theory of markets whether conventional or unconventional (S. Winter). How firms in oligopolistic markets take potential entry into consideration when setting prices or building capacity will have important implications for both short-run and long-run behavior and performance. Given the important role of entry in many of the formal models, as well as in the oral tradition of industrial organization, it is very surprising that there have been such limited efforts directed at building a richer understanding of who this character, the "potential entrant," might be, what the nature of the entry decision process is, the nature of the interdependence *between* potential entrants (R. Sherman and T. Willet), and how existing firms perceive potential entry and react to it.

Instead of assuming that some simple type of interrelationship exists among firms, more effort might be directed at both theoretical and empirical modeling of pricing and entry decision processes themselves, in different kinds of markets including the implications of organizational structure, the nature of the product and technology, the history of the industry, and other important institutional characteristics. The effort may not only imply new mathematical and analytical approaches, but also require one to leave the office and library to spend some time with actual firms that make actual pricing decisions.

Can one indeed pursue a program such as this and come up with more than just a catalogue of descriptions? I believe that it can indeed be done, but it requires an analysis of the ways in which decision-making processes evolve in response to various phenomena which characterize the real world environment in which firms operate. Contact with actual firm decision-making processes can be helpful in suggesting the kinds of phenomena which influence the decision-making structure and the nature of alternative processes. More formal analysis and empirical work can seek to provide simple structural models of these internal processes and analyze the ways in which individual firm behavior adds up to observed market behavior in both the short run and the long run.

A number of important phenomena appear to have implications with regard to pricing and entry decisions as well as other important decision-making processes. Let me suggest what these phenomena are and discuss each briefly.

A. Uncertainty

Firms operate in a dynamic world characterized by a great deal of uncertainty. For example, a firm will normally expect to have some random variation in demand from one period to the next. At a particular output and price level its inventory policy serves as a buffer between transitory fluctuations in orders and its production schedule. However, one is always uncertain whether a particular change in orders represents a random fluctuation or a secular shift in demand. If it is the latter, both short-run and long-run changes in prices and production plans may be called for. There is some evidence that firms have fairly simple inventory rules which trigger adaptive behavior by the organization. However, the type of behavior that is observed appears to differ from one industry to another. In some industries the immediate response triggered by a perception of a permanent fall in demand is a price reduction (R. Cyert and J. March, 1963). In other industries the response may be a decrease in production (or ordering for a retail firm) but the maintenance of existing prices (automobiles). The nature of such differential responses no doubt depends on characteristics of the product and technology as well as inventory holding costs, but is also heavily influenced by considerations of bounded rationality, contracting (and recontracting) costs, organizational structure of the firm, and market ethos. As a result we can expect to observe differences in response to changes in demand, input prices, taxes, etc. from one industry to the next *given* a particular level of a traditional market structure index such as the concentration ratio.

B. Bounded Rationality

It has long been recognized by psychologists and some economists that human beings are not capable of nor inclined toward continuous global optimization given all readily available information (H. Simon). Rather, decision makers opt for relatively simple rules of thumb to guide decisions. In addition, the status quo has certain stability characteristics arising from satisficing rather than maximizing behavior. Since it is expected that oligopolistic markets will be relatively more insulated from competitive pressures than are perfectly competitive markets, one might expect the existence of rule of thumb decision making and the sluggish response to change arising from satisficing behavior to be even more prevalent in such markets.

In addition, given a market ethos characterized by cooperation rather than by hard-nosed rivalry, simple rules of thumb may be desirable as a means of coordinating such cooperation.

Recent research by R. Radner has shown that considerations of bounded rationality can in fact be cast in a rigorous analytical framework yielding interesting behavioral propositions. Radner's "firefighting" manager may be relatively impervious to changes in particular economic circumstances if the effects are not large enough to reach the top of the manager's list. Among other things, firms exhibiting this kind of managerial behavior may not respond quickly to changes in certain costs or to the entry of new rivals if larger "fires" exist which will be fought first. If such behavior exists, it may have important implications for the effectiveness of alternative public policy instruments—for example, the use of Pigouvian pollution taxes vs. the use of mandatory rules and controls. To the extent that such rule of thumb behavior is more prevalent in insulated oligopoly markets, it might help to explain relative price rigidities and give us a deeper understanding of the importance of entry, as well as to indicate the desirability of applying different public policy instruments to sectors with different market structures.

C. *Information and Contracting Costs*

The existence of information and contracting costs have important implications for many aspects of firm behavior. The oligopoly firm is not only concerned about information regarding the prevailing market price, input costs, etc., but is also concerned about what rival firms are doing and what potential entrants might do. The firm's ability to monitor and respond to the actions of rivals, both real and potential, is limited by considerations of both bounded rationality and information costs. Its ability to form, monitor, and enforce informal collusive agreements is determined by the same considerations. The smaller are information and contracting costs, the greater will be the ability of a small group of firms to behave the way one single monopoly firm behaves. The larger are information costs, the less likely will a firm be able to identify and react to the threat of potential entrants. Similarly, the more costly is information about actual prices, costs, and profits achieved by existing firms, the smaller the set of potential entrants and the less likely will be actual entry from this set of firms, other things equal.

The notion of "proximity" is important, especially with regard to information about prices, entry, and technology. While by proximity I mean the closeness of a particular firm to others, I do not conceive of it in a geographical sense. A firm may be close to another firm because they produce the same product, use similar technologies, or are in contact with one another via market exchanges of products. A firm may be closer to another firm to which it sells and/or buys products than it is to a firm in its own market, especially with regard to actual transactions prices. Firms with similar technologies may be closer to one another than firms which produce complementary products. Firms borrowing money from the same financial institutions may be close to one another even though they produce different products and use different technologies. The topology of the terrain on which the firm operates is determined by much more than a product dimension. It is determined also by a technology and transactions dimension. To the extent that a firm has search processes regarding products, prices, costs, profits, and technologies, it will look most intensely at those firms which are near it because information costs are low. The more densely packed is the space surrounding the firm (across all dimensions), the lower are the information costs per bit of information, but the more the information that must be collected, the higher the contracting and enforcement costs if coordination of behavior is desired. Oligopolistic coordination would be difficult in such a market. Alternatively, the firm might have only a small number of existing firms producing the same product which are in close proximity. Information costs per bit of information from these firms would be low, as would total information costs and contracting costs. It would be costly to obtain information about other firms not in the product market, but these other firms would face the same problems. This situation facilitates the persistence of simple decision-making rules and monopolistic coordination.

Considerations of information and contracting costs have not been used extensively in the analysis of firm and market behavior. Nelson and Winter include both rules of thumb and the notion of proximity in an evolutionary model of economic growth and technological change. Their simulation approach facilitates an analysis of the adding up characteristics of a market composed of firms with particular behavioral rules. O. Williamson has applied the idea of contracting costs as part of his institutional failures approach to internal organization. In a simple satisficing model of pricing decisions of regulated firms, I have tried to specify how information about the unobservable allowed rate of return is obtained from observations of other firms. An extension of these approaches to deal explicitly with the oligopoly problems would be fruitful.

D. *Organizational Structure*

A firm's organizational structure is itself partially determined by considerations of uncertainty, bounded rationality, transactions costs, and technology. The complete characterization of the internal organization of the firm should include not only measures of hierarchical complexity, but also information gathering and processing mechanisms and decision-making structures. But since many large firms produce several products and decision-making processes evolve over time, once the basic organization is formed it is worthwhile to examine regularities that link stimulus with response via particular organizational structures. Attempts to relate characteristics of the organizational structure to decision making in the oligopoly situation have been quite limited (Cyert and March, 1955 and 1956), but are nevertheless of great interest. This is especially true with regard to antitrust policy, since it can influence organizational structure in certain circumstances. Organizations with complex hierarchies and centralized decision making are likely to respond more slowly and more routinely to changes in costs, demand, and actions of rivals and potential entrants than would a less hierarchical and more decentralized decision-making structure. But the centralized decision-making unit may be better able to integrate different kinds of information about the firm's activities and environment and to pursue carefully coordinated policies regarding pricing, production, advertising, etc. To the extent that some decisions (for example, pricing) are made at a different level of the hierarchy than are others (let's say production decisions), we may observe what appear to be biases in the responses to changes in the economic environment (like price rigidity but output flexibility). It would be interesting to see how much of the relative price rigidity often associated with concentration is really determined by differences in organizational structure. To the extent that organizational structure influences the nature and speed of a firm's responses, further analysis along these lines may also be useful in designing effective antitrust remedies.

E. Market Ethos

Perhaps the most difficult thing to get a handle on is what I call the ethos of the market. Some industries are characterized by social and historical relationships which favor cooperative behavior rather than hard-nosed rivalry. Behavior which has been characterized by "conscious parallelism" or "tacit coordination" is obviously facilitated in a market in which firms both want to coordinate and trust one another. I suspect that such a cooperative ethos is most prevalent in fairly old industries which are growing fairly slowly and where firms are close to one another in the sense used above. Rapidly expanding industries selling new products and declining industries are less likely to exhibit the cooperative ethos. Entry of new capital as well as new owners (via merger or takeover) will also affect the ethos of the market. Market ethos and internal organizational ethos (A. Hirschman) are certainly closely related. It is tempting to dismiss this phenomenon as being outside the domain of economics and hand it to sociologists and psychologists. I do so only reluctantly, because I have the feeling that market ethos is an important determinant of market behavior in small numbers situations. To the extent that other social scientists do have something worthwhile to say about this phenomenon, I hope that we will make use of it in our policy analysis and policy prescriptions for particular markets and not simply ignore it.

IV. Conclusion

Pricing and entry behavior are only two of several areas where a deeper understanding of decision-making processes and how they evolve in different market contexts is extremely important. Research and development and finance are prime candidates. Space doesn't really permit me to go into them, however, and the first has been discussed quite nicely elsewhere (Nelson 1972). But since I've trampled over many of the flowers in the garden, let me step on a few more in conclusion.

The sick sister of the structure-conduct-performance nexus of industrial organization is the conduct part or, as I would rather refer to it, behavior. The oral tradition of industrial organization abounds with many behavioral stories about pricing, entry, research and development, financing, etc., but they have not been incorporated into our theoretical conception of industrial markets. We have spent too much time calculating too many kinds of concentration ratios and running too many regressions of these against profit figures of questionable validity. We have spent too little time analyzing regularities in behavior that arise from structural characteristics other than concentration, including the structure of internal organization, the nature of basic learning, diffusion, and information processes, the closeness of products and technologies, the role of financial institutions, and yes indeed, considerations of business psychology and sociology. We haven't done it because the traditional tools of the theory of the firm are not very helpful. Nevertheless, the essence of the oligopoly problem is behavior, and if the tools are not to be found in the conventional theory of the firm, we might consider developing tools that are appropriate. Hopefully, the industrial organization area will in this way play an important leadership role in the extension and revision of the conventional theory of the firm rather than be its prisoner.

REFERENCES

J. Bhagwati, "Oligopoly Theory, Entry Prevention and Growth," *Oxford Econ. Pap.*, Nov. 1970, *22*, 297–310.

D. Baron, "Limit Pricing, Potential Entry, and Barriers to Entry," *Amer. Econ. Rev.*, Sept. 1973, *63*, 666–74.

A. Chandler, *Strategy and Structure*, Cambridge 1962.

R. Cyert and J. March, "Organizational Structure and Pricing Behavior in an Oligopolistic Market," *Amer. Econ. Rev.*, Mar. 1955, *45*, 129–39.

——— and ———, "Organizational Factors in the Theory of Oligopoly," *Quart. J. Econ.*, Feb. 1966, 44–64.

——— and ———, *A Behavioral Theory of the Firm*, Englewood Cliffs 1963.

J. Dirlam and A. Kahn, *Fair Competition: The Law and Economics of Antitrust Policy*, Ithaca 1954.

W. Fellner, *Competition Among the Few*, New York 1949.

F. Fisher, "New Developments on the Oligopoly Front: Cournot and the Bain-Sylos Analysis," *J. Polit. Econ.*, Aug. 1959, *67*, 410–13.

C. Frank, Jr., "Entry in a Cournot Market," *Rev. Econ. Stud.*, July 1965, 245–50.

D. Gaskins, "Dynamic Limit Pricing: Optimal Pricing Under Threat of Entry," *J. Econ. Theory*, Sept. 1971, *3*, 306–22.

A. Hirschman, *Exit, Voice and Loyalty*, Cambridge 1970.

P. Joskow, "Pricing Decisions of Regulated Firms: A Behavioral Approach," *Bell J. Econ.*, Spring 1973, *4*, 118–40.

M. Kamien and N. Schwartz, "Limit Pricing and Uncertain Entry," *Econometrica*, May 1971, *39*, 441–54.

A. Kaplan, J. Dirlam, and R. Lanzillotti, *Pricing in Big Business*, Washington 1958.

F. Modigliani, "New Developments on the Oligopoly Front," *J. Polit. Econ.*, June 1958, *66*, 215–32.

R. Nelson, "Issues and Suggestions for the Study of Industrial Organization in a Regime of Rapid Technical Change," in *Policy Issues and Research Opportunities in Industrial Organization*, New York 1972.

———, "The Link Between Science and Invention: The Case of the Transistor," in *The Rate and Direction of Inventive Activity*, Princeton 1962.

——— and S. Winter, "Neoclassical Versus Evolutionary Theories of Economics Growth; Critique and Prospectus," *Econ. J.*, forthcoming.

M. Piore, "On the Technological Foundations of Economic Dualism," Mass. Instit. of Tech., dis. pap. 110.

W. Primeaux, Jr., and M. Bomball, "A Reexamination of the Kinky Oligopoly Demand Curve," *J. Polit. Econ.*, July/Aug. 1974, *82*, 845–62.

R. Radner, "Some Stochastic Models of Administrative Allocation," mimeo., Aug. 1974.

F. Scherer, *Industrial Market Structure and Economic Performance*, Chicago 1970.

H. Simon, "Theories of Decision-making in Economics and Behavioral Science," *Amer. Econ. Rev.*, June 1959, *49*, 253–83.

J. Simon, "A Further Test of the Kinky Oligopoly Demand Curve," *Amer. Econ. Rev.*, Dec. 1969, *59*, 971–75.

R. Sherman and R. Tollison, "Technology, Profit, and Market Performance," *Quar. J. Econ.*, Aug. 1972, *86*, 448-62.

——— and T. Willet, "Potential Entrants Discourage Entry," *J. Polit. Econ.*, Aug. 1967, *75*, 400–03.

G. Stigler, "The Kinky Oligopoly Demand Curve and Rigid Prices," *J. Polit. Econ.*, Oct. 1947, *55*, 432–49.

——— and J. Kindahl, *The Behavior of Industrial Prices*, New York 1970.

O. Williamson, *Markets and Hierarchies: Analysis and Antitrust Implications*, New York 1975.

S. Winter, "Satisficing, Selection and the Innovating Remnant," *Quar. J. Econ.*, May 1971, *85*, 237–61.

Oligopoly Theory, Communication, and Information

By MARTIN SHUBIK*

The stage is set for the development of a mathematical institutional economics which is simultaneously more abstract and more institutional than much of economic theory to date. The need for this development can be seen clearly by examining both the successes and failures of oligopoly theory.

Oligopoly theory has been the skeleton in every economic theorist's closet. Virtually every course in theory has a few hours spent in telling the students about the problems posed and showing them a few classical "solutions," after which the subject is dropped. The reason for this treatment is because we do not have a generally accepted adequate theory of oligopoly.

The reasons why we do not have a satisfactory theory are central to many economic problems. They exist for microeconomic models of competitive markets and they are hidden in the assumptions made in macroeconomic models. However, it is possible to gloss over the basic difficulties in these because, although they are still there, they do not appear to play as central a role as they do in oligopoly theory.

The key items are:

(1) the number and "size" of economic agents,
(2) the communication system,
(3) the state of information,
(4) the specifics of structure of both industrial and financial institutions, and
(5) the roles and goals of economic actors.

These items primarily involve the description of the structure of a firm, industry, or market. Considerations of structure (especially in reference to information and communication) cannot easily be separated from considerations of behavior. Thus we may need to reconsider the nature of the decision making of the individual we choose to regard as the economic decision maker.

I. Different Approaches to Oligopoly Theory

There are four broad approaches to oligopoly theory which merit distinguishing. Not only are they complementary with each other, it is my belief that in the new theory that is emerging they will come together. The approaches are:

(1) mathematical models of oligopolistic competition,
(2) institutional studies and industrial organization,
(3) "neoclassical" oligopoly theory, and
(4) behavioral models of oligopoly, the "new industrial organization" and gaming experimentation.

Ever since Cournot, oligopoly theory has provided a rich source of models for the mathematically inclined. The works of A. A. Cournot, J. Bertrand, F. Y. Edgeworth, H. Hotelling, T. Bowley and others attest to this as do the more recent works using game theory including R. J. Aumann, R. Selten, Shubik, B. Shitovitz, L. Telser, and others.

Most of the mathematical models have

---

* Yale University. The research was supported by the Office of Naval Research and also partially funded by a grant from the Ford Foundation.

been static. They have divided more or less naturally into those stressing some variant of a noncooperative equilibrium as a solution and those considering cooperative solutions such as a contract curve or the core.

The research and teaching on industrial organization has scarcely intersected with the mathematical work, although this is now beginning to change. Nevertheless, until recently the painstaking detailed studies of individual industries and firms, such as those of A. F. Kahn and M. G. DeChazeau, J. Peck, F. M. Scherer, J. B. Dirlam, and others, have been rarely if ever read by those concerned with mathematical models. Being institutionally oriented, this work may be read by lawyers before other economists. (See, for example, H. Packer.)

Neoclassical oligopoly theory is scarcely well defined, but I would lump under this title E. H. Chamberlin, W. Fellner, H. Brems, F. Machlup, G. Stigler, H. von Stackelberg, P. Sylos-Labini, J. J. Bain, and several others. They are all less mathematical than the first group and less institutional than the second. Chamberlin's contribution, for example, was mathematically less rigorous than that of Cournot, but contained much more relevant economic modeling. Bain's contribution on the study of entry is closer to an institutional study than the others in this group.

The key characterizing features of what I have called "neoclassical oligopoly theory" are that the analysis is in an open or partial equilibrium context; much of the discussion is cast in dynamic or quasi-dynamic terms; information implicitly plays a role; money explicitly plays a role (trade is in goods for money and vice versa) and the solution sought is some sort of noncooperative or quasi-cooperative equilibrium.

In general, in this type of theorizing (with some exceptions, Bain and Brems, for instance) there is little in the way of micro detail on the differences in the communication and information patterns in different industries or on critical technological factors which limit the scope and timing of moves.

This type of oligopoly theorizing is sufficiently spiritually close to the style of Marshall's partial equilibrium analysis and comparative statics that it supplies the major source for the two or three lectures on oligopoly that are thrown into the microeconomics courses.

The fourth approach to oligopoly theory has come about through a basic dissatisfaction with the other three. The key characterizing feature of this approach is a central concern with the explicit description of process. It is explicitly concerned with dynamics and with disequilibrium. There are possibly three different modes of development which can be discerned. They are:

(1) behavioral models of the firm and simulations,
(2) experimental oligopoly investigations, and
(3) mathematical models of search and other economic activity under incomplete information.

These approaches cannot be easily compartmentalized as they overlap. The first is best characterized by the work of R. M. Cyert and J. G. March, R. Nelson, S. Winters, and several others. It is spiritually linked to the large business games, such as the Carnegie Tech game, as well as to operations research work in marketing, production scheduling, finance, and other special functions.

An important link among economic theory, operations research, and computer simulation of behavior comes in the attempts which have been made to build models of the behavior of specific firms or parts of firms (see A. M. Schrieber). Much of this work is done without academic blessing; many studies done by and for private firms do not see the light of academic day and probably many do not deserve to be taken too seriously. Nevertheless, the efforts expended in attempting to construct detailed simulations are beginning to provide a connection between institutional and theoretical studies. The simulations are frequently rich in detail, yet at the same time spell out process.

Experimental oligopoly research is small relative to other scholarly activities, but since 1948 it has been growing and includes work by Chamberlin, T. Dolbear, L. Fouraker, J. Friedman, A. Hoggatt, Selten, Shubik, S. Siegel, V. Smith, and others.[1] There are several messages that this work has to deliver, the major one being that the specific details of communication, information, and the mechanisms of the market have considerable influence on the play when numbers are few.

Most of the experiments have employed students or individuals who are not senior businessmen; the time of play has been relatively brief and the stakes have been for the most part small or nonexistent. Thus what we are learning about is the behavior of students playing in business or economics games with low financial incentives and little if any institutional identification. Even so several useful observations can be made. Numbers play an important role. Without face-to-face communication the numbers in an experimental game appear to be 1, 2, 3–6, and "many." Information overload and computational limitations are also clear in the behavior of business game players. The coding and information processing step going from a matrix game to one with balance sheets is large. In all of the work the importance of sociopsychological variables is evident. These include "bargaining ability" in the work of Siegel and Fouraker, "trust" and "cooperativeness" in the work of Hoggart, T. C. Liu, and others. Furthermore, different approaches to cooperation and to decision making are evident in the work of Siegel and Fouraker, D. Stern, Shubik and G. Wolf, and others.

The third of the new approaches covers explicit models of search and behavior where information is scarce and costly. This includes work by P. Diamond, H. Leland, J. McCall, D. McFadden, M. Sobel, M. Spence, and others.[2]

II. A Reconciliation of General Equilibrium and Oligopoly Disequilibrium

General equilibrium theory and partial equilibrium offer us the concept of the efficient anonymous market price system. Oligopoly theories, in contrast, suggest the noncooperative equilibrium: quasi-cooperative solutions involving leaders and followers, "kinked oligopoly reaction functions," cooperative agreements, or a variety of behavioral mechanisms.

When general equilibrium theory is contrasted with work in oligopoly theory, it appears to be so pristine, general, mathematical, and noninstitutional that it makes the latter appear in utter disarray. This could be because oligopoly theorizing is not abstract enough or because there may be within the presentation of the general equilibrium system a false generality which conceals many gaps in basic modeling. I believe the latter is true.

There are at least eight desiderata for any robust economic model of markets which any student of oligopoly will recognize. A reasonable model should:

(a) depend explicitly on the number of participants;
(b) contain an explicit description of the functioning of the market mechanism;
(c) be able to handle nonsymmetric in-

---

[1] For a detailed bibliography, see M. Shubik.

[2] A recent article by M. Rothschild provides a survey.

formation conditions;

(d) be defined for all positions of disequilibrium as well as equilibrium;

(e) be able to handle markets with few traders and other markets with many traders together;

(f) have assets and capital structure play an important role;

(g) have money and finance play an important role; and

(h) be consistent with behavioral models of economic process.

The general equilibrium model as presented by K. J. Arrow, Gerard Debreu, and others fails to satisfy any of these criteria. The noncooperative behavioral model first suggested by Cournot as a solution to oligopolistic competition can be modified so that it satisfies all eight.[3] The work of Chamberlin, Stackelberg, Fellner, and many others and the recent work in game theory as applied to economics can be interpreted as variations on this theme.

The variety and difficulties manifested in oligopoly theory are a reflection of the maze of institutions and complex of processes which make up economic life. There is an underlying structure to these economic processes but in order to appreciate it we must be simultaneously more abstract and more concrete or institutional than most of current microeconomic theory. The theory of games, the new work on search and decision making under uncertainty are beginning to provide the basis for the former; process models, simulations, and direct empirical work are providing the latter.

REFERENCES

H. Packer, *The State of Research in Antitrust Law*, Walter E. Meyer Research Institute of Law, 1963.

M. Rothschild, "Models of Market Organization with Imperfect Information," *J. of Pol. Econ.*, 1973, *81*, 1283–1308.

A. M. Schrieber, ed., *Corporate Simulation Models*, Seattle 1970.

M. Shubik, *The Uses and Methods of Gaming*, New York 1975, ch. 10.

———, "A Theory of Money and Financial Institutions, Part XVI: Mathematical Models for a Theory of Money and Financial Institutions," C.F.D.P. No. 377, to appear in the *Proc. of the Symposium on Adaptive Econ.*, Math. Res. Center, Univ. of Wisconsin.

---

[3] For a discussion of this somewhat cryptic remark, see Shubik.

EMERGING STRUCTURAL CHANGES IN TRANSPORT AND PUBLIC UTILITIES

Northeast Railroads: Restructured or Nationalized?

By ALEXANDER LYALL MORTON*

The intractability of the current Northeast rail problem owes to the fact that the demands of shippers, the demands of rail employees, the claims of creditors, and the demands of public transport policy cannot all be met by the Northeast rail system—they add up to more than the rail system, operating in the competitive environment it does, can produce. Even if the claims of the creditors on the bankrupt estates were to be extinguished, the present demands of the other three claimants could not be met by a self-supporting rail system. Briefly stated, the situation is thus: various labor practices on the railroads (by no means the fault solely of rail labor unions) raise rail costs to the point that most freight traffic (and virtually all passenger traffic) has been lost to other modes of transport and to the point that railways are only marginal competitors for most remaining freight traffic. Large rail shippers use their muscle, playing railroads off one against another and against competing modes to hold down rail freight rates and to extract services from the carriers at prices below cost. Government policy imposes a host of common carrier obligations on the railroads, requiring "essential public services" to be performed and cross-subsidized by profits on other services. Intense competition from other modes and the power of large shippers have eroded such profits (which once derived from markets in which railroads enjoyed monopoly powers). The owners of the Northeast railroads have for some decades been the victim of this squeeze play, but now their resources have been largely exhausted, and the Northeast rail plant is in shambles. To whom do employees, shippers, and the beneficiaries of public policies turn next to fill the gap created by their excessive demands upon the rail system?

The political process has intervened in the search for a solution. Declaring the bankrupt railroads to be of concern to the national interest, Congress passed the Regional Rail Reorganization Act of 1973, thereby removing the solution from the exclusive jurisdiction of the bankruptcy courts. In fact, the political process did not have to intervene; conventional thinking has for many years exaggerated the importance of the railroads to the Northeast economy (as we will suggest below). There is, of course, a risk to the government's taking on this responsibility: Will the political process be able to force compromise and sacrifice upon the competing interest groups, or will it ultimately resort to the public purse to squeeze past the crisis?

The proper remedy is to force the claimants to moderate their demands so as to create a viable, self-supporting rail system, for there is no good reason why freight transport should be a subsidized factor of production. Happily, the medicine need not be too bitter, but it does require some rather radical departure from past practice.

---

* Harvard Business School.

First, let us look at the changes labor must accept. Restrictive labor practices raise the labor cost of rail freight transportation to roughly equal the labor input of private and exempt motor carriage, thereby contributing to the loss of rail traffic to trucks. Work rules, craft lines, seniority agreements, and a host of related restrictions severely limit and constrain the specific tasks that may be assigned to any individual rail employee. The purpose of these rules is to share the "available work" more widely, thereby protecting jobs and income. But these understandable goals of labor can be served in a much better way. Slow, unreliable train service and the rundown condition of track and equipment reveal that there are plenty of opportunities for using extra useful service from every present employee to improve service and upgrade plant. To guarantee that greater labor productivity is not turned against labor, employment and income guarantees should be issued. In fact, improved rail service would generate rail freight traffic, actually improving employment prospects in the future. Greatly improved rail labor productivity is one essential remedy for the Northeast rail problem.

Let us turn now to the reform required of shippers and public transport policy. Shippers, local communities, and passengers exploit the common carrier obligations and regulation of the railroads to obtain and preserve rail services at prices below their true cost. The bankruptcies of Northeast railroads were caused to a large degree by forcing the carriers to provide many services at unremunerative rates, and the restoration of these carriers to profitable operation requires an end to this practice.

The proper way to do this is to release railroads from common carrier obligations altogether. Free railroads to negotiate detailed service contracts with individual shippers that specify what services will be provided and what price will be charged, and let the railroads abandon all services whose cost the market refuses to bear, just as most other private industry is permitted to do.

To understand why such a radical cure is necessary and, at the same time, feasible or safe, one must look at the intense competition Northeast railroads now face. American industry has a variety of transport modes to choose among to move its freight; the railroads receive only about one out of every eight dollars spent on intercity freight movement. Equally important, industry has a rich variety of alternatives to the purchase of transportation; among the most important of these are decentralization of production and substitution of locally available materials for those which must be moved longer distances. Thus rails are but one alternative in an array of logistical options that industry has. Let us focus more intently on the truck as a competitor of and alternative to rail. The cost functions of rail and private and exempt truck operations are today nearly identical. This is perhaps astonishing because of the apparent broad differences between the two technologies, but it is nonetheless true. Any commodity can move by truck and move at a cost that is, at best, only slightly higher than the current cost (to the railroad) of rail movement. Trucks can and do carry large amounts of virtually every commodity there is to be moved. (Coal is the commodity whose movement railroads most dominate; yet trucks move about one-sixth as much coal as rail and their share is growing.) Virtually every shipper and every community has unhindered access to trucking for its freight needs. Modern highways serve virtually every location; there is no legal restriction on entry into private trucking; and the near total absence of economies of scale associated with truckload operations means that even the comparatively small shipper whose freight needs will fully utilize only a truck or two has access to virtually all the benefits of truck transport. Because of its broad availability, cost competitiveness with rail, and generally better service and greater convenience, trucking has grown to be the dominant mode. Trucks presently receive five times the revenues of railroads for intercity freight movements.

Trucks, then, can and do compete on a cost basis for virtually every type of traffic the railroads carry, and the cost of private and exempt trucking places a very firm and reasonable ceiling on the price an unregulated railroad industry could possibly demand for its service. The one area where railroads can, one may argue, develop distinct cost advantages over trucks is in high-volume, long-distance movements of bulk commodities such as coal, ore, petroleum products, chemicals, construction aggregates, lumber, and grain. Industries that rely on these raw materials have come increasingly to locate along the expanding network of navigable waterways, however, and water carriers, either alone or in conjunction with truck, now offer cost-competitive and surprisingly widespread competition to the railroads for that one type of traffic that may be somewhat insulated from truck competition. Also, production and distribution of these bulk commodities have become highly concentrated in the hands of a few, very large firms. The size and market power of these firms more than protect them from abusive treatment by railroads.

Shippers no longer require common carrier obligations and regulation to protect them from the railroads. Rather, freedom from these obligations may even be necessary to the survival of railroads in the Northeast. Competing modes are closing in on rail freight transport. They have already diverted vast amounts of rail traffic in the Northeast and they are becoming ever more cost and service competitive for the rest. The point is being approached where the volume of rail traffic remaining in the Northeast fails to achieve the economies of scale and density necessary to support the Northeast rail system in anywhere near the form we know. The grip of the Northeast railroads on the freight market is so tenuous and uncertain that, far from being restrained, they must be given the freedom and encouraged to compete aggressively for all the traffic they can get.

Given this freedom, the railroads may once again carve out for themselves a defensible position in the Northeast. The opportunity is there. The dedicated, or unit, train—a train composed of cars designed specially for their load that shuttles back and forth between specified origin(s) and destination(s)—permits substantial savings in labor, fuel, and capital costs over conventional train operations. It should be used more widely for the movement of bulk commodities and those manufactures that generate repetitive, high-volume movements. Piggybacking—the movement of truck trailers on railway flatcars—enables the railroads to recapture other traffic that has been lost to trucks. The railroad's cost advantage over trucking lies primarily in line-haul efficiency. Piggybacking permits railroads to capture the line-haul portion of intercity freight movements while leaving to trucks the local pickup and delivery functions that trucks clearly do best. Northeast railroads have been moving toward greater use of unit trains and piggybacking, but the common carrier obligations imposed on railroads severely retard what should amount to a technological revolution, a massive restructuring of the industry.

Having intervened in the Northeast rail problem, is the political process facing up to the hard decisions outlined above? The signs are not encouraging. The Regional Rail Reorganization Act itself virtually foreclosed the possibility of substantial reform of restrictive labor practices. The one lever the authors of the Act possessed to goad management and labor into work-rule revision was the offer of protection for any employees that might be affected in rationalizing the Northeast rail system. The Act, however, grants this protection, on extraordinarily generous terms, without requiring any meaningful change of labor practices.

It is still too early to say what success the Act and the instrumentalities it creates—the United States Railway Association (*USRA*) and the Consolidated Railway Corporation—will have in ending the exploitation of the common carrier obligations imposed by the Interstate Commerce Act. While some pruning of rail operations in the Northeast has been spoken of, hearings held throughout the region have revealed strong political opposition to service cutbacks of any kind. The representation of shippers, states, cities, rail labor, and the Interstate Commerce Commission on the Board of Directors of *USRA* suggests that meaningful reform on this front can be little more than a hope.

There are, on the contrary, disturbing signs that we are being prepared for an assault on the public purse to get us past this crisis. Any employee whose job is eliminated will be generously protected by public funds, according to the Act, though there are few if any precedents for public protection of workers who have lost jobs when the fortunes of an industry have turned. Branch line subsidies will preserve "essential rail service" along rail lines that might otherwise be abandoned, though the decline of traffic on these lines is itself evidence that rail service is no longer essential to more than a few shippers. Fixed rail plant is to be rebuilt with federal grants and loan guarantees.

One can even hear the rhetoric that is to prepare us for continuing subsidy of Northeast rail operating deficits after these initial excuses—labor protection, branch-line subsidies, etc.—begin to pall. The railways, we are told, are a "valuable natural resource," the "low-cost carrier," a "fuel-conserving mode of transport." Is it cynical to ask why, if railroads are so inherently efficient, have they sunk to their present state and why must they be subsidized and protected from higher-cost competitors? Judging from present deficits, the drain of the Northeast rail system on public funds will run on the order of $500 million to $1 billion annually even at the onset.

But the present drift of events will almost certainly carry us further than nationalization of Northeast railroads. Clever, solvent rail carriers may find it profitable to encourage federal involvement and public subsidy for the bankrupt railroads. Freight cars today absorb over half of all rail capital expenditures. The Penn Central, western and southern carriers claim, has long beggared other railroads for freight cars because of its own inadequate ownership. What an opportunity to reverse the score now that the public treasury will be buying cars. The argument for branch-line subsidies does not lose validity at the Potomac, Ohio, and Mississippi Rivers; once the principle is accepted for the Northeast, solvent carriers should be able to tap easily into a branch-line subsidy program.

Then look further to the future. The entire American railroad industry has earned a subnormal return on invested capital for decades. Net railway operating income in 1973, a good year, was one-third lower than in 1929, without adjustment for the greatly reduced purchasing power of those dollars. The railroads' rate of return on transportation investment has not exceeded 5 percent since 1929 and has not exceeded 4 percent since 1955. Railroad executives constantly face a difficult decision in committing and recommitting capital to an enterprise whose return has been so marginal for decades, particularly now that interest rates have risen to the 10 percent level. With the low returns and the vast appetite for capital that railroads have, the owners may welcome a chance to unload the business, if only some sucker may be found to buy it. The settlement with the creditors of the Northeast bankrupts may set the terms at which the owners of solvent railroads can expect to sell out. While the Regional Rail Reorganization Act may not have intended it, it appears that the courts may award the estate reasonable value for any properties taken over. Once management of solvent carriers decides that liquidation of rail assets and ultimate sale of the remains to the government are more attractive than perpetual reinvestment for inadequate returns, the properties of these carriers will begin the long slide into disrepair, down the same track as the Northeast bankrupts.

Other industries may some day find it attractive to play the same game. Once the government commits itself to solving industrial crises like the Northeast rail situation and doing so from the public purse, an important restraint upon the demands of other interest groups in the economy will have been sacrificed. Already several large airlines have applied, though unsuccessfully to date, for Federal subsidies. Distant early warnings of trouble in the electric utility industry have also appeared. Is it time to ask what precedents we wish to set?

Transportation Controls and the Spatial Structure of Urban Areas

By GERALD S. GOLDSTEIN AND LEON N. MOSES*

The nineteenth century city in America relied upon rail transportation, the horse-drawn streetcar, as the primary mode of travel to get people to and from work. In the twentieth century city the automobile replaced the streetcar as the dominant mode for commuters. Traffic congestion, air pollution, noise pollution, urban blight, and "wasteful" use of gasoline have been attributed to the growing use of automobiles, particularly for the work trip. Transportation policies that would encourage the use of public mass transportation relative to the private automobile have been proposed by public interest groups and government agencies, such as the Environmental Protection Agency, in order to alleviate some or all of these problems.

The analysis in this paper will focus on the long-run effects of transportation policies that affect the cost of automobile travel versus mass transit on the spatial distribution of households. In the long run households can adapt to such controls by moving their residences, and/or their places of work. A basic urban land-use model developed by urban economists will be adapted to explore the long-run changing patterns of land use. Inferences must be drawn from a partial equilibrium approach, since there does not exist a static model of the city in which the intraurban location of firms and households is simultaneously determined.

Formal economic models of the city have focused predominantly on the spatial distribution of households. In these models all production and employment are located in a central point through which all imports and exports move. This point is assumed to be located on a featureless plain on which transportation is ubiquitous. The population is composed of households with identical utility functions who bid for land around the central point, the central business district (*CBD*). The cost of building and maintaining houses (excluding site rent) is constant throughout the city.

In equilibrium all households must achieve the same utility level. Since all jobs are located in the *CBD*, people who live some given distance from the *CBD* must pay higher costs of transportation to get to the *CBD* than people who are located nearer to the *CBD*. Because of the equilibrium condition, and the assumption that transportation costs are positive and increase with distance from the *CBD*, those households who live near the *CBD* must pay higher rent per unit than those households living farther away. In equilibrium, the price per unit of housing must decline with distance from the *CBD* so that what the household saves in housing costs by moving further away from the *CBD* must just be offset by the increase in costs of getting to the *CBD* as a result of the move. A further implication of this model is that population density will decline with distance from the *CBD*, the rate of decline depending on the rent gradient.

---

* Department of Economics and Center for Urban Affairs, Northwestern University, and Department of Economics, Center for Urban Affairs, and Director of Transportation Center, Northwestern University, respectively. Research support from the Motor Vehicle Manufacturers Association is gratefully acknowledged. The views expressed are ours and not necessarily those of the association.

The higher is site rent at any given location, the higher population density will be.

There is one modification to this basic model which is of interest here. We introduce the additional assumption of local employment opportunities uniformly distributed throughout the area. The *CBD* remains the dominant source of employment opportunities. The implication of this modification is that there will be a wage gradient, analogous to the rent gradient. The basic explanation of this wage gradient is the same as that for the rent gradient. Because household equilibrium requires equalization of household utility throughout the area, people who live and work outside the *CBD* must earn less than people who live at the same location but work in the *CBD*. If they both received the same wage, then the worker working outside the *CBD* would earn a higher real income since he would not incur the transportation costs of the *CBD*-worker, but would pay the same price for all other goods.

We turn now to the long-run effects of transportation controls on urban spatial patterns. We must consider the influence of alternative modes of transportation on wage, rent, and price gradients. This issue must be faced since a variety of transportation controls are designed to increase the cost of auto relative to public transportation. Further, a new dimension, choice of work area, must be added. We will assume two employment centers, the core and a suburban center. The wage rate for a particular kind of skill is assumed to be higher in the core than in the secondary center. A very dense system of public transportation, perhaps bus, is assumed to exist around the core. Workers have three choices: (1) to live within the bus commuting area and use the bus to commute to the core; (2) to live outside the bus commuting area, but commute to the core for work by car; (3) to commute to the secondary center by car.[1] The situation is depicted in Figure 1 where O is the core, and A, the secondary employment center.

The solid circle around the core is the limit of the bus commuting area. Those who live on the boundary are indifferent between going by bus to the core and going by car; those inside the circle commute by bus, and those outside it commute by car. The circle represents the results of a situation in which in a sense there is complete coverage of the inner area by bus routes and the bus is less expensive in money terms than the car, but requires more time. In addition the downtown parking fee is greater than the round-trip flat bus fare.[2]

The car commuting area is divided into two zones by the hyperbolic market boundary UU'. Those who live on UU' are indifferent between traveling to the core for work and traveling to the secondary center for work. They are indifferent because the real income they achieve is the same regardless of the choice they make. Individuals who live inside the boundary UU' commute to the secondary labor market by car.

The analysis begins with all households in equilibrium defined above to mean that no household can improve its level of satisfaction by changing its place of work, residential site, or mode of transportation. In addition, all land markets are cleared. A change is now made in the cost of transportation. Two broad alternatives are considered. The cost of travel by car is increased or the cost of public transportation is reduced. These changes can be brought about in a variety of ways. The impacts of most of them on the wage and rent gradients differ only in detail. However, an in-

---

[1] The model has been constructed to eliminate choices that would add complexities, but which would add little to the problem being investigated. As constructed, the problem eliminates cross-hauling and what is often called back-commuting, i.e., commuting to the secondary center by people who live close to the core.

[2] If the per mile cost, including the value of time, is less by car than bus, there is no bus area.

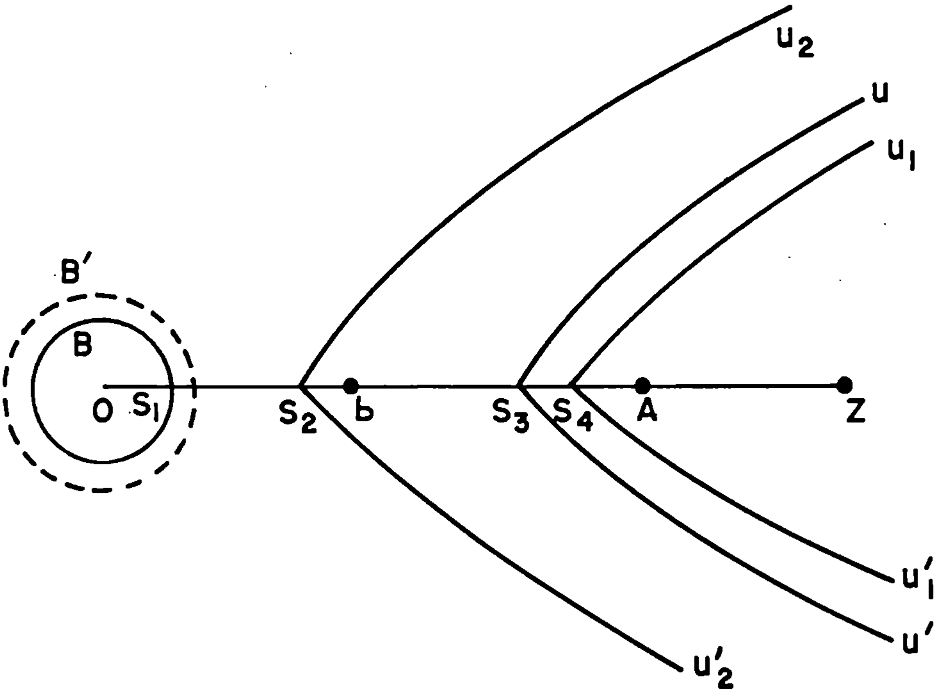


Figure 1

crease in downtown parking charges or in tolls does have somewhat different implications from an increase in the per mile cost of auto travel. Therefore these measures must be studied separately.[3] The effects of different measures that improve the attractiveness of public transportation also differ somewhat, but here the differences are insignificant.

Overall, the procedure followed is to change the quality of the public transportation system and examine its effects on the wage gradient and the household bid-rent function. Implications of the public transportation improvement for the new equilibrium distribution of population and economic activity are then sought. The same line of reasoning is then worked out for an increase in per mile auto costs and the effects are compared.

The quality and cost characteristics of public transportation can be improved in a variety of ways. There is no need to consider all of them since they appear to have very similar effects. We deal with a case in which an express bus is introduced whose time and cost characteristics make it an attractive alternative to the car for a certain set of households. The initial impact of the new service on choice of mode is represented in Figure 1. The express bus service is assumed to operate in the automobile commuting area. So far as the analytical model is concerned, it does not matter where the station is located. It can be somewhere on a line different from *OZ*, in the secondary center itself, or on the line *OZ* and between *O* and *A*. The latter is assumed. *b* is the station.

The introduction of express bus service makes it cheaper in time, money, or disutility of travel to commute to the core by bus. It thereby raises the net wage that can be earned by working in the core. This causes the boundary UU' to shift to the position shown by U_1U_1'. In the short run households that live in the area between these two boundaries continue to work in the secondary employment and commute by car. In the long run they may change place of employment and place of work. Initially it is assumed that they are satisfied with their residential area and seek to adjust to the change in transportation by changing place of employment.[4] The effect of the new service is to raise the labor supply function of the core and reduce the labor supply function of the secondary center and any other actual or potential employment centers in the automobile commuting area. The statement made above concerning the two labor supply functions can be phrased differently. The new service causes the individuals who live between UU' and U_1U_1' to search for jobs in the core. It also increases the wage rate that they would demand at the secondary center and at all places within this area between UU' and U_1U_1'.

---

[3] The effect of these measures and their impact on decentralized cities as well as on monocentric cities is studied in greater detail in Goldstein and Moses, "Air Pollution, the Clean Air Act, and the Spatial Structure of Urban Areas," mimeo. 1974.

The express bus also opens up a second choice for those who have been commuting to the core by car. This choice is represented by U_2U_2'. Individuals who live to the right of this boundary and to the left of UU' find that their net incomes are higher if they shift from car to the new service. If they are satisfied with their neighborhoods, the effect of the improvement is to increase the wage rate that they would demand for working nearer home, or anywhere in the satellite area as an alternative to continuing to work in the core. This means that the wage gradient shifts up in the satellite area between UU' and U_2U_2', which is to say that there has also been a downward shift in the labor supply function of any actual or potential employment site in this part of the satellite area.

Households who live to the left of U_2U_2' and outside the bus circle continue to travel to the core by car. Nevertheless, their costs of travel may be affected by the new service. If their costs change, their choice between working in the core and working nearer home or at other potential employment sites in the area will be affected. Suppose that the introduction of the new bus service reduced automobile congestion in and around the core because some workers have changed to bus. This will reduce the cost of automobile commuting to the core for workers who live to the left of U_2U_2' and make labor supply conditions less favorable in the automobile commuting area. In other words, the minimum wage individuals will accept for employment in their part of the satellite area is increased.

Finally we consider the individuals who live in the local bus circle. They are not directly affected by the new service. The reduction in automobile commuting to the core from other areas might increase the speed and convenience of local service, but this is unlikely to be a strong effect. The wage gradient in this area is therefore treated as unchanged.[5]

Over the largest part of the former automobile commuting area, labor supply conditions for local employment have become less favorable. In addition, land in the area has become more valuable for households who commute to the core for work. These two changes may make the area less desirable as a location for those kinds of industry, such as manufacturing, where single floor layout is valuable and large amounts of space are used. It seems likely that an improvement in public transportation would act to make the automobile commuting area a less favorable location for many kinds of industry. It would certainly not contribute to a more rapid rate of business migration from the central city. However, as population expands in the area around the station, retail as well as some service activity, such as insurance and banking, might be expected to grow.

---

[4] This assumption obviously alters the traditional model of urban location. In effect we are introducing neighborhood effects and saying that access to work is not the sole determinant of residential location.

[5] The conclusions reached would not be altered significantly if local service did improve as a result of the introduction of express service. In this case, the households who live in the bus area of the central city are less willing to work nearer home and outside the *CBD*.

The area in the vicinity of the secondary employment center has become relatively less valuable for households. This means that industry could acquire additional land at less cost. Industry has some incentive to expand in the area. On the other hand, labor supply conditions have become less favorable. It is not possible to judge the relative strength of the two changes. However, if the secondary center does not have advantages such as access to low-cost interregional transport facilities, it is unlikely to grow rapidly as an employment center. The effects of an increase in the cost of automobile travel will now be considered.

This cost can be increased in a variety of ways. Consider first the effects of an increase in the per mile costs of such travel, not bothering to distinguish between an increase due to a transportation measure or a market increase in the price of gasoline. There has been a tendency to view increases in automobile travel costs as having the same implications as improvements in public transportation. In the short run both have the effect of causing diversion from car to public transportation. However, the long-run effects on the distribution of population and economic activity are likely to be quite differferent. Improvements in public transportation are likely to slow the rate of outmigration of industry from core areas and central cities generally. Increases in the cost of auto travel are very likely to cause core areas and central cities to decline more rapidly.

An increase in the per mile cost of auto travel causes a shift from car to local bus. This is represented as an increase in the bus zone from B to B', in Figure 1. Individuals who live between the two circles have been made worse off, the amount of their reduction in real income increasing with distance from the core. These individuals are more willing than before to work at sites closer to their homes and more accessible by car. This means that the labor supply function of the core has decreased and the labor supply functions of actual and potential employment sites in the zone between the two circles have increased.

The increase in auto travel cost also shifts in the boundary UU' to the position U_1U_1'. In the short run the individuals who live between the two boundaries cannot change their place of residence or employment. However, they have suffered reductions in their real incomes. The farther they live from the core the greater is the amount of the reduction. If we treat residences as fixed for the moment, it is clear the individuals who live between the two boundaries are more willing to work at sites that are nearer home. This means that they are willing to accept lower wage rates than before for employment at such sites. Labor supply conditions then are less favorable in the core because its labor supply function has shifted down.

The individuals who live to the right of UU' continue to work in the secondary center, but they have also suffered a reduction in real income. The further they live from the secondary center, the greater the amount of the reduction. They would also like to work closer to home.

Overall, if households react to an increase in auto travel costs by searching for jobs from their given residential sites, the effect is to improve labor supply conditions in suburban residential areas. This area includes that segment where individuals have shifted to bus. The effect of the increase is to make labor supply conditions in the central city less favorable. This change in labor supply conditions will clearly increase the relative attractiveness of noncore sites for industry.

Some additional comments are necessary. Households who live in the original bus area, an area corresponding roughly to the central city, have not experienced a change in real income while all other households are worse off. In this situation it can be argued that city living has improved relatively, and that the improvement will cause individuals to return to the city. We do not believe that such a remigration will occur. Hence, we are saying that such readjustments as take place in residential settlement will not change the generally adverse impact that an increase in auto travel costs is likely to have on labor supply conditions in the central city. Consider now the possible implications of an increase in auto travel costs on the value of space.

An increase in the per mile cost of auto travel reduces the net money income of households over the entire automobile commuting area. Overall, the effect of such a reduction should be to decrease what households can bid for residential space and thereby make space cheaper for industry. In other words, taking the automobile commuting area as a whole, the increase in travel costs is likely to have an impact on land values that reinforces its effect on labor supply conditions. The attractiveness of the satellite area for industry is increased still further. The above is not meant to preclude the possibility that there will be changes in relative land values over the satellite area. We cannot judge this, again because it is not possible to say anything definite about the new long-run equilibrium and the pattern of jobs and households that will emerge in the satellite area. If individuals' commuting costs go up substantially, but truck transport costs do not increase very much, we are likely to see a broad dispersal of economic activity. There would be a tendency for many small employment centers to emerge, each drawing labor from a relatively localized area. However, if truck costs rise substantially, there will be a tendency for larger employment centers to develop, particularly in those portions of the satellite area that have good access to low cost interregional transport facilities.

I. Summary and Conclusions

This paper has examined the long-run effects on cities of changes in the cost of moving people. Two kinds of policies are compared: (1) those that increase the cost of travel by car, and (2) those that improve the quality or reduce the price of public transportation. The paper argues that the former are likely to reduce auto travel more quickly. However, these policies are also more likely to cause the economies of central cities to decline more rapidly than they would otherwise. A policy of improving and subsidizing public transportation would tend to slow their rate of decline. So far as the economies of central cities are concerned, there is a set of measures that appears to be most damaging. These are the ones that increase the cost of auto travel to central cities or their core areas relative to the cost of auto travel elsewhere in standard metropolitan statistical areas. Reductions in the supply of downtown parking, special taxes on such parking, tolls to the downtown area, etc. are examples of such measures. Unfortunately, since funds for improvements in public transportation almost certainly must come from taxes, there may be an inclination to rely on such measures to achieve air quality in large central cities.

Regulation and the Financial Condition of the Electric Power Companies in the 1970's

By Paul L. Joskow and Paul W. MacAvoy*

Electricity accounts for about 25 percent of the total energy consumed in the United States and its share of total consumption has been increasing slowly. In an effort to decrease the vulnerability of the United States to foreign energy price increases and embargoes, Federal energy policy, particularly as formulated in the *Federal Energy Administration Report for Project Independence*, would accelerate this trend towards electricity. These policies call for the mandatory conversion of household and commercial heating to electricity and the conversion of oil- and gas-burning plants to domestically available coal and uranium. Even without such mandatory controls, consumers may choose more electricity as a result of increases in fuel oil prices relative to electricity prices, or as a result of the shortage of natural gas, or because electricity supply seems more secure.

Although Federal policies and consumer choice may shift demands towards electricity, there is no assurance that the additional quantities and mix of generating capacity desired will in fact be forthcoming. The nation's investor-owned utilities (providing over 90 percent of generating capacity) are not likely to be able to raise the required amounts of capital. Increases in construction costs, fuel costs, and interest charges have recently outstripped revenue growth, and expectations that this trend will continue have made utility investments unattractive. The suspicion is that regulatory procedures have recently caused price increases to lag behind cost increases, resulting in earned rates of return below the cost of capital. If this continues, capacity to meet increased demands—and Project Independence, in whatever form—will not be achieved.

The purpose of this paper is to assess the financial prospects of the nation's electric utility industry, given existing regulatory institutions and continued high rates of growth of demand in the late 1970's. Shortages from regulation would indeed be a turn of events. Economists' analyses of regulatory effects in the 1960's deplored the behavior of commissions on the grounds that they did nothing to control prices (see G. J. Stigler and C. Friedland and R. Jackson). However, the turn of events would not be entirely surprising; using the behavioral approach in analyzing regulation leads one to suspect that commissions operate relatively independent of economic conditions (see Joskow, 1972). The well-established operating rules of the bureaucracies change only when the results of such procedures under inflation or depression become intolerable (see Joskow, 1974). What may have been ineffective regulation in the 1960's may be overzealous regulation for the late 1970's.

I. State Regulation of Electric Utility Prices

Retail prices for electricity, in keeping with long-established mandates from most state legislatures, are regulated by state public utility commissions. The commissions are supposed to monitor the financial performance of firms under their jurisdiction, increasing or decreasing prices so as to allow a return on capital that is "fair." In practice, state commissions appear to be much more passive than this, conducting formal reviews only when firms themselves seek to increase prices. Other procedures, including Dear John letters or jawboning, are used to reduce prices when rates of return rise above what was allowed in the formal case. But the process itself indicates more commission resistance to price increases than initiatives for significant price decreases (see Joskow, 1974).

---

* Assistant professor of economics and Henry R. Luce Professor of Public Policy, Massachusetts Institute of Technology, respectively.

The price increase review process, when initiated, has been highly formalized to center on historical accounting data. First, a test year is chosen for analysis of accounting costs and revenues at existing prices. Operating costs, depreciation, interest charges, and general and Federal taxes (minus any tax credits), plus the allowed profit, are calculated for test-year production. The total of these amounts sets the limit on total earnings. The critical calculations are of the allowed profits—equal to the allowed rate of return times the rate base (the value of capital equipment used in the production of the good or service under regulation). The whole procedure is essentially backward looking; the earned rate of return in the test year is compared with the allowed rate of return, and prices are adjusted upwards or downwards to the point where the rate of return which would have been earned during the test year equals that allowed. Projections or use of future test years have not been used extensively and are now only being considered by regulatory commissions in a few states.

The procedures take a great deal of time. A formal rate case for one of the larger utilities requires first the administrative law judges and then the commissions to examine *ex parte* materials and collections of testimony on all aspects of costs, profits, and prices. This may take from ten months to one year. By allowing temporary rate increases, this processing lag may not affect price immediately, but given that the commissions then are not pressed by waiting companies, a final decision is sometimes made based on data that is over two years old.

The commissions not only take time, but they follow rather simple rules of thumb that do not often require much of the information collected over the review period. In a study of the New York Public Service Commission, it has been found that, on the critical determination of the rate of return, what the commission allows depends upon (a) capital costs, (b) the size of the requested increase in total revenues, and (c) the presence of objecting intervenors (see Joskow, 1972). The last two determinants involve large elements of judgment and bulk large in the results since the commission did not use an economic concept of the cost of capital to set limits on allowed rates of return. Capital costs have more likely been defined as average historical rates of return of other utilities. As these costs held constant or fell, regulation eased; but as they increased under inflation, the invoking of old rules led to more restrictive regulation. The New York Commission in the last few years has judgmentally allowed a larger percentage of requested revenues, specifically in response to inflation. Otherwise, judgments center on extrapolating past judgments (see Joskow, 1972, pp. 643–44). The question is whether this New York response to inflation has been picked up by other commissions and has been sufficiently flexible to keep the rate of return above the cost of capital.

The procedure itself does not warrant optimism. With sharply increasing construction costs and doubling of interest rates, future costs and historical test-year costs bear no relationship to each other.

Revenue erosion can well result where earned rates of return will fall below allowed rates of return (and perhaps below the costs of capital). There are additional sources of flexibility, however. One way of circumventing the established decision rules is through the use of automatic adjustment mechanisms that allow prices to increase or decrease as certain incurred costs change without resorting to a formal regulatory review. Such a mechanism has been widely adopted for fuel costs as a portion of operating expenses: as fuel prices increase, electricity rates are allowed to increase by the same amount on a per kilowatt basis. Thus the regulatory process essentially sets an adjustment formula which then operates automatically as that cost component changes.[1]

Such flexibility, whatever the benefits, has been purchased at potentially high economic costs. The automatic passthrough of fuel charges, along with the slower and more cumbersome regulatory review of other costs, can lead to distortions. The choice of generating techniques may therefore be biased towards energy-intensive turbines away from capital-intensive nuclear capacity. This is to assert the existence of a negative Averch-Johnson effect when waiting costs and rates of return below marginal costs of capital from regulation are entered into the decision-making processes of electric power firms.

II. Costs, Prices, and Financial Performance, 1964–74

Given this regulatory structure, it is instructive to look at costs, prices, and the financial performance of the investor-owned electric utilities in the United States during the past ten years. From 1964 until 1968, operating expenses were essentially constant at about sixty-five cents per kilowatt hour, but thereafter began to rise rapidly and had reached a level of ninety-six cents per kilowatt hour in 1973. At the same time, prices fell during a good part of the decade. From 1964 to 1969, average price per kilowatt hour fell slightly, but beginning in 1970 has risen steadily with the largest one-year increase occurring in 1974 (primarily as a result of increased fuel prices triggering fuel adjustment clauses in most states). The rise in prices came about as a result of increases in the number of formal regulatory reviews from an average of four per year during 1964–68 to fifty-three in 1972 and fifty-six in 1973.

The financial condition of the industry was not greatly improved by the price increases. Interest charges, depreciation, and profits account for the margin between revenues and operating costs. Interest charges, which in 1964 were 4.5 percent on new public utility debt, had risen by 1970 to the level of 8.8 percent. With less stringent monetary policy in 1971 and 1972, the rates decreased, but they have risen since then to 9.7 percent in June 1974. Given such increases, profit rates have declined. The rate of return on equity peaked in 1967 at 12.8 percent and has fallen every year since then (except one) to 11.2 percent in 1973 and 10.9 percent for the first nine months of 1974.[2] Thus the difference between the rate of return on equity and the yield on new long-term debt, which was nearly eight percentage points in 1964, fell to a little more than one percentage point in 1974.[3]

---

[1] Without the added flexibility of automatic fuel adjustment, a number of utilities, especially in the Northeast and middle Atlantic regions, would have been faced with liquidity crises in the face of fuel oil price changes during the winter of 1973–74. But such clauses are being challenged in the courts in a number of states and they may not survive as part of the regulatory process throughout the 1970's.

[2] It should also be noted that in real terms (adjusted by the consumer price index) earnings per share began to fall in 1967 and by June 1974 were below what they had been in 1964. Finally, the interest coverage level for the industry fell continuously over the period to a value of 2.1 by September 1974, extremely close to the limit in many mortgage bond indentures.

[3] The quality of these earnings has also been subject to some doubt. Allowances for "construction work in progress," a noncash item, have recently appeared as a

This financial performance has led to investor expectations that future profits will yield returns below the cost of capital. Stockholders have indicated that they believe this to be the case: by June 1974, average electric utility stock prices had fallen to about 75 percent of book value. External financing is difficult to achieve under these conditions—falling earnings and high interest charges soon put the firm up against legal limits on issuing new debt, and a continued policy of selling stocks at high yields but substantially below book value reduces earnings per share at a multiplicative rate. Thus there may well be limits on the ability to raise enough capital to expand generating capacity.

III. Future Performance under Various Regulatory Regimes

This discouraging outlook could be changed by the behavior of the commissions. The state agencies learned to adjust to rising costs in the late 1960's, at least to the extent that they did not so closely regulate prices that new capacity was difficult to finance. If the commissions allow new inflation premiums, they could in effect provide profits sufficient to raise enough capital to expand capacity at the old rate. Given much higher construction outlay and much greater capital costs for interest and equity financing, these premiums could be spectacular in size.

An initial investigation of the question of premiums is undertaken here through a small-scale experimental financial model of an electric power company. The model is designed to provide some insight into the implications of various regulatory policies under various states of the economy. It does not forecast what the commissions are going to do, but rather restricts discussion to the implications of alternative commission approaches to inflation and rapid capacity expansion. A flow chart of the model appears in Figure 1.

In the model, total revenue is calculated by multiplying the price by the number of kwhs sold for each consuming sector (residential, commercial, and industrial). Each expense item is calculated in detail in accordance with the accounting procedures generally used by regulatory commissions. New investment requirements are met by the use of retained earnings and external financing via long-term debt, preferred stock, and common equity. The company can sell long-term bonds at a specified interest rate as long as its debt ratio is less than 70 percent and its interest coverage is greater than 1.65. It can sell new equity and preferred stock as long as it has earnings after the payment of dividends to existing stockholders sufficient to pay equivalent dividends to new stockholders. If earnings are insufficient to allow for the sale of new debt, or preferred stock, and/or equity (which in total equal external financing) then the State Power Authority is required to provide the additional assets at a specified interest rate. This residual claim represents the capital shortage that the utility faces, and it seems realistic to have it made up by a state entity as investments that have to be made to prevent power outage. The figure therefore gives a measure of potential capacity shortages that might result under different regulatory regimes.

Choosing a sample of twelve of the larger flow-through electric power companies, we use the model to examine four regulatory scenarios. In each case we examine the earnings per share, rate of return on equity, and the ability of the firms to

---

source of earnings of increasing importance. In 1964 this item accounted for only 5 percent of total earnings, but by 1974 it had increased to almost 35 percent of reported earnings. Since this is not a current cash source of income, but rather a credit which will eventually be capitalized in the rate base (when the construction is completed), the earnings figures do not give a good indication of the utility's actual cash flow position. Thus earnings per share and, in some cases, dividends have exceeded cash flow per share by significant amounts.

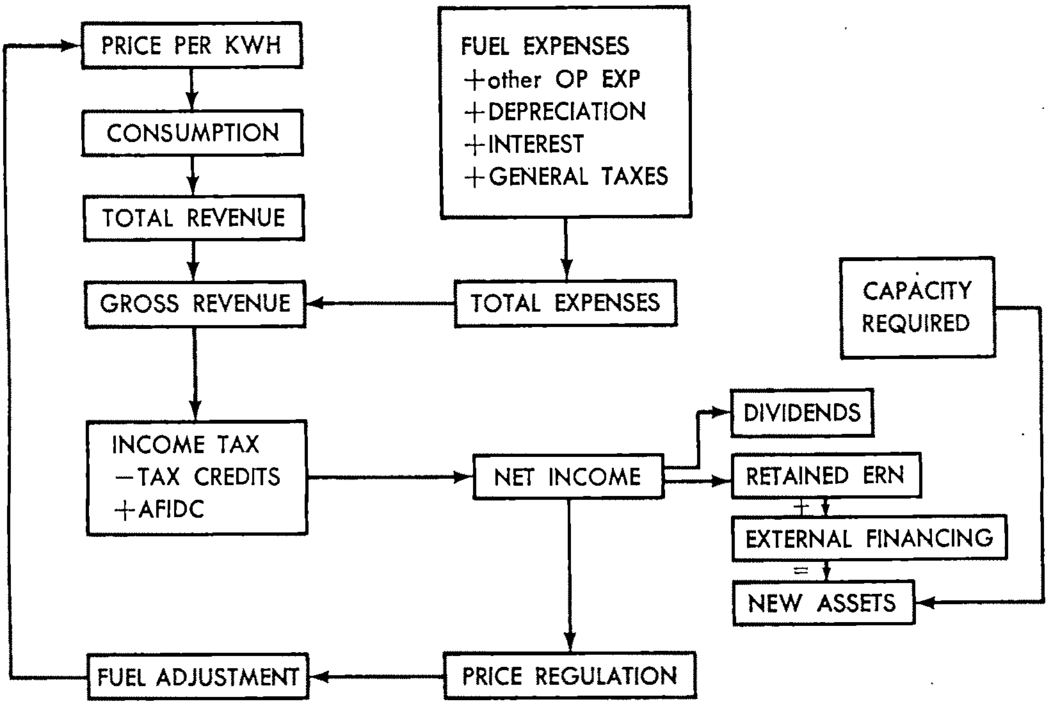


FIGURE 1: FLOW CHART FOR ELECTRIC UTILITY FINANCIAL MODEL

raise the required amounts of capital using retained earnings and external financing for the period 1975–80. Each scenario assumes that Project Independence is a reality, so that capacity has to at least continue to grow at the 1969–72 trend rate. All consumption is assumed to grow at two-thirds of the 1969–72 trend rate. Fuel price changes are assumed to pass through at 100 percent and it is assumed that all 1974 fuel price increases are maintained and increase at a 6 percent rate during the period.

Case 1: Rate of return on equity target of 10 percent used in regulatory hearing plus complete fuel adjustment. Price adjustments are based on historical test-year data (latest twelve months).

Case 2: Same as Case 1 except target is 12 percent on equity.

Case 3: Same as Case 2 except target is 14 percent on equity.

Case 4: Same as Case 3 except target is 16 percent on equity.

In each case we assume that long-term interest rates and preferred stock yields remain high at 10 percent. If equity can be sold, it is at the mean 1973 price/earnings ratio for each company.

Table 1 reports the results for Case 2. Ten of the twelve companies face serious financial problems under these conditions. The capital shortage figures show that these companies will have to either defer some capital acquisitions or rely on state subsidies to obtain the needed capital.[4]

---

[4] To the extent that companies merely defer further expansion, the actual earnings figures reported by the companies will be higher than those projected by the model (because of the absence of the State Power Authority). This may be especially true for the 1974 predictions since one year's needed growth can easily be deferred without serious consequences. In the long run, however, these higher earnings are achieved at the expense of reduced system reliability and perhaps actual shortages by the early 1980's.

TABLE 1—FORECAST EARNINGS PER SHARE, RATE OF RETURN ON EQUITY, AND CAPITAL SHORTAGES WITH A TARGET RETURN OF 12 PERCENT ON EQUITY

Company	Earnings per Share			Rate of Return on Equity			Capital Shortage (Annual in $ millions)	
	1973	1974	1980	1973	1974	1980	1974	1980
American Electric Power	2.9	2.1	2.3	12.4	9.9	11.1	463.1	880.8
Cincinnati Gas & Electric	2.3	1.9	2.2	13.5	10.5	12.2	0.0	112.8
Consolidated Edison (N.Y.)	2.3	2.0	2.3	6.7	6.0	7.6	758.0	1,102.4
General Public Utilities	2.3	1.8	2.4	10.7	8.1	10.9	101.6	593.6
N.Y. State Electric & Gas	3.0	3.0	5.0	9.9	9.8	12.3	0.0	0.0
Northeast Utilities	1.4	1.1	1.4	11.7	8.8	11.2	173.0	311.9
Ohio Edison Company	2.4	1.6	1.7	14.9	10.3	11.3	114.2	410.1
Pacific Gas & Electric	3.2	1.9	2.6	10.8	6.8	10.1	0.0	788.6
Public Service Elec. & Gas	2.2	1.2	2.3	9.1	4.4	8.6	594.5	590.1
Southern Calif. Edison	2.7	2.2	4.9	10.0	7.6	11.4	0.0	0.0
Union Electric Company	1.6	1.4	2.0	10.7	8.9	12.3	4.0	106.4
Virginia Electric Power	2.1	1.3	1.7	11.2	6.9	10.3	428.5	778.9

Table 2 presents the projections of the financial model for 1980 for all four cases. In most cases, the earned rate of return on equity is below that allowed. This result emerges because of the use of historical cost, revenue, and balance sheet data to set returns in the regulatory process while these data do not accurately reflect the behavior of these items in the year that the rates are actually in effect. At 10 percent allowed rate of return on equity, only Southern California Edison and New York State Electric and Gas face no capital shortage. They are able to achieve this result at a fairly low rate of return because they are growing slowly and can finance all new construction through the sale of long-term debt and preferred stock plus retained earnings. Most of the other companies exhibit very substantial capital shortages relative to their capital needs at this allowed rate of return.

When we turn to the 16 percent rate of return on equity case, we see that the

TABLE 2—PERFORMANCE IN 1980 UNDER FOUR REGULATORY REGIMES

Company	CASE 1		CASE 2		CASE 3		CASE 4	
	ROR[a]	Shortage (Annual in $ millions)	ROR	Shortage (Annual in $ millions)	ROR	Shortage (Annual in $ millions)	ROR	Shortage (Annual in $ millions)
American Elec. Power	8.4	886.4	11.1	880.8	13.1	888.4	15.3	163.6
Cincinnati Gas & Elec.	10.2	113.8	12.2	112.8	14.4	112.2	14.9	0.0
Consolidated Edison (N.Y.)	4.8	1,107.2	7.6	1,102.4	10.6	1,090.9	11.5	257.6
General Public Utilities	8.4	591.0	10.9	593.6	12.7	94.7	14.0	0.0
N.Y. State Elec. & Gas	10.5	0.0	12.3	0.0	13.8	0.0	15.2	0.0
Northeast Utilities	8.9	303.7	11.2	311.9	12.7	4.2	14.5	0.0
Ohio Edison Company	8.8	416.7	11.3	410.1	12.4	259.2	13.3	0.0
Pacific Gas & Elec.	7.6	799.3	10.1	788.6	10.0	61.9	11.5	0.0
Public Service Elec. & Gas	6.1	591.4	8.6	590.1	12.2	583.1	13.2	194.9
Southern Calif. Edison	9.5	0.0	11.4	0.0	13.2	0.0	14.7	0.0
Union Electric Company	9.9	128.0	12.3	106.4	12.9	0.0	15.0	0.0
Virginia Electric Power	7.2	780.2	10.3	778.9	13.9	723.4	13.2	0.0

[a] ROR = Rate of return on common equity actually earned.

situation has changed dramatically. Nine of the twelve companies exhibit no capital shortages. In addition, while the three remaining companies still exhibit capital shortages at the 16 percent rate of return, they are only a small fraction of those incurred at 10 percent.

IV. Conclusions

Although the financial model used here is experimental and is in an early stage of development, several conclusions emerge which deserve consideration.

1. The allowed rate of return on equity of 10 or 12 percent chosen by some regulatory commissions will lead to severe financial difficulties and associated capacity shortages.
2. Allowed rates of return of at least 14 percent appear to be necessary to generate sufficient profits to attract financing for capacity growth in line with Project Independence.
3. The critical value of the allowed rate of return on equity which will lift a company out of financial difficulties varies from one company to the next, depending on the rate of growth in output, capacity, and operating expenses. These differences arise in part because of the historical cost-based nature of the rate-setting process which itself leads to earned rates of return below those allowed. This revenue erosion effect seems to be a pervasive characteristic of existing regulatory institutions.

All things considered, the maintenance of existing regulatory procedures may lead to financial disaster for the electric utility industry. Without substantial increases in the rates of return on equity allowed by regulatory commissions, or without changes in the historical cost rate-making process itself, it is unlikely that sufficient capital can be raised to achieve historical growth rates in capacity. Since commissions seem reluctant to base rates on equity returns as high as 14 or 16 percent, the answer must lie in more creative use of future test years, automatic adjustment mechanisms, and alternative methods of calculating a company's rate base. These are, however, only short-run measures which treat current symptoms rather than the root causes of the problem. Efforts must be directed towards incorporating considerations of the marginal cost of production in the rate-setting process. Such a procedure would by necessity be forward looking and would allow prices to increase or decrease as the marginal cost of producing changes over time.

REFERENCES

R. Jackson, "Regulation of Electric Utility Rate Levels," *Land Econ.*, Aug. 1969, *45*, 372–76.

P. L. Joskow, "The Determination of the Allowed Rate of Return in a Formal Regulatory Proceedings," *Bell J. of Econ. and Management Science*, Aug. 1972, *3*, 632–45.

———, "Inflation and Environmental Concern: Structural Change in the Process of Public Utility Price Regulation," *J. of Law and Econ.*, Oct. 1974, *17*.

G. J. Stigler and C. Friedland, "What Can Regulators Regulate?," *J. of Law and Econ.*, Oct. 1962, *5*, 1–16.

Development Economics—A Reassessment of Goals

By IRMA ADELMAN*

The theme of this paper is the need to be more specific on the goals of economic development and the interdependence between values and goals. This theme is not original; Gunnar Myrdal has argued for many years the importance of values in determining the orientation of development economics. Implicitly, by concentrating on growth, development economists have assumed that other goals, such as greater equity, need not be considered because they are positively correlated with growth. There has been some discontent with this view, but convincing scientific evidence refuting it has not become available until recently. Given that the relationships among the goals of distribution, welfare, and development are by no means simple, it is clearly necessary to examine the purposes of development and to state explicitly the value premises underlying the basic goals. In this paper I shall first review some recent work which bears on these relationships. I shall then propose a minimum humanistic goal for development and, finally, shall suggest development strategies that may permit achievement of that goal.

I. The Evidence

In this section I shall draw primarily on two recent efforts: my latest book with Cynthia Taft Morris and the results of comparative statics experiments performed with a detailed, instrument-specific, general equilibrium model of the South Korean economy, constructed by Sherman Robinson and me.

A. A Review of the Adelman-Morris Findings

Using 1950–63 cross-section data for forty-three underdeveloped countries, Adelman and Morris studied the relationships between the shares of income accruing to the poorest 60 percent of households on the one hand and various broad aspects of a nation's economic, social, and political performance on the other. We found that for the longest part of the development process—corresponding to the transition from the state of development of sub-Saharan Africa to that of the least developed Latin American countries—the primary impact of economic development on income distribution is, on the average, to decrease both the absolute and the relative incomes of the poor. Not only is there no automatic trickle-down of the benefits of development; on the contrary, the development process leads typically to a trickle-up in favor of the middle classes and the rich. The absolute incomes of the poor begin to rise with development only when the nation moves well into the "intermediate" level of development. Further, even here improvement is not automatic: "the poorest segments of the population typically benefit from economic growth only when the government plays an important economic role and when widespread efforts are made to improve the human resource base . . . " (Adelman-Morris, p. 181).

---

* Professor of economics, University of Maryland. With the usual disclaimer of their responsibility, I am indebted to F. L. Adelman, J. L. Joy, C. Taft Morris, S. Robinson, and P. P. Streeten for their most helpful comments.

These results imply that, for a lengthy portion of the process of economic development, there is a conflict between the growth of overall national income and an increase in the welfare of the poor. Equally important, there is an analogous conflict with respect to policy instruments for systematic intervention.

It follows then that, in the formulation of development planning models and development policy, it is necessary to decide in advance the extent to which increases in the welfare of the poor are to be weighed against simple growth of *GNP*. In addition, traditional economic instruments of policy appear to have only a weak, or at most a nonsystematic, effect on the relative share of income accruing to the poor. Indeed, our study "reinforces the view that the policy instruments that are most effective in improving income distributions are different from those that are best for raising economic growth rates" (p. 185). It also supports the view that "economic structure, not level or rate of economic growth, is the basic determinant of patterns of income distribution" (p. 186).

Implicit in these conclusions is the suggestion that the price of economic equity is high: a necessary condition for its achievement is radical structural change.

B. *The Adelman-Robinson Model*

In order to test the validity of this last hypothesis, an income distribution-oriented planning model of an ambitious and novel variety was formulated.[1] The results of comparative statics experiments with the model reinforce the conclusions of the Adelman-Morris book. Dynamic experiments are now in process.

The model is designed to investigate the impact of a variety of traditional economic instruments on income distribution: relative factor prices, product prices (influenced directly through price fixing, taxes, subsidies, and/or world prices, and degree of monopoly), rationing of products and services, rationing of credit, government taxes and transfers, trade policy, labor market policy, technology, and inflation. The model is dynamic and instrument-specific. It captures the effect of various policy instruments on distribution through a modified general equilibrium approach in which both factor prices and product prices are endogenous. Given the initial ownership of assets and the institutional rules of the game, the short-run income distribution derives from the interaction of production and consumption decisions.

The model is capable of portraying a large variety of economic and institutional rules of the game. One can vary the degree of monopoly, the principles of operation of credit markets, the clearing principles for labor markets and for commodity markets, and the permissible degrees of disequilibrium in individual markets. The model combines Keynesian and Walrasian analysis (it can operate in either mode, depending on the labor market specification). It portrays a monetary economy (combining Schumpeter-Wicksell with Walras) open to foreign trade. It can be run in an equilibrium or in a disequilibrium mode. The model is rich in sectoral detail (twenty-nine sectors), is disaggregated by firm size (four per sector), and has fifteen factors of production and twelve socioeconomic consumer (= saver) household categories.

Solution for each time step is by tatonnement in a two-stage process. In the first stage an allocation of investible funds is made across sectors and firm sizes by making firm demand (based on production functions and anticipated sales, rates of return, and capital cost) consistent with a given interest rate, a target credit expansion, or a direct rationing program. The second stage of the model is a microeconomic market model rich in institutional, sectoral, firm size, and household category detail. A series of markets must clear—a process that determines production, employment, prices, wages, and income distribution. Before the next period the model updates the variables, sets expectations, and selects credit and trade regimes for the first stage of the next period.

---

[1] Work on this model has been supported by the International Bank for Reconstruction and Development.

This model has been implemented for South Korea. Some indication of the validity of the model may be inferred from the fact that the base period solution for the model reproduces the 1968 data to within less than 1 percent for every one of the approximately 2,000 endogenous variables.

Some of the comparative statics experiments performed to date are listed in Table 1. The results of these experiments are summarized in Tables 2 and 3. Each column of Table 2 indicates the distribution of household income after the model is solved under the conditions of the corresponding experiment, indicated at the top of the column. There appears to be a remarkable insensitivity of relative income distribution to the experiments performed. Table 3 indicates the composition by socioeconomic group of the lowest decile for each of the experiments. The bottom row of the table gives the average real household income of the decile. An examination of this table suggests, by contrast, that economic policy, by changing the functional distribution of income, can play a major role in determining who are the poor

TABLE 1—DESCRIPTION OF COMPARATIVE STATICS EXPERIMENTS
Adelman-Robinson Model

Experiment	
A–1	Peg agricultural prices at 1.25 of base price.
A–2	Increase agricultural capital stock by 50 percent.
A–3	Increase agricultural productivity 25 percent in cereals and 15 percent in other agriculture. Peg both prices at .90 of base price, and allow free trade in agricultural products.
A–4	Increase productivity in cereals by 25 percent and give cereals a price subsidy of 25 percent.
B–1	Double all direct taxes on households and transfer the proceeds to skilled labor, apprentices, and workers.
B–2	Give price subsidy of 20 percent to labor intensive sectors. Increase their capital stock by 20 percent. Peg their prices to world prices, allowing unlimited exports.
B–3	Same as B–2 but, in addition, increase agricultural productivity by 25 percent.
C–1	Increase entire capital stock by 50 percent.
C–2	Increase capital stocks by 50 percent for nonagricultural firms with less than 50 employees (two smallest sizes).
D–1	Increase all exports by 25 percent.
D–2	Devalue by 30 percent. Peg prices of traded goods to world prices. Increase capital stocks of trading sectors by 20 percent.
E–1	Increase supply of engineers, technicians, skilled workers, and clericals by 10 percent, reducing the number of unskilled workers correspondingly. *CES* functions.
E–2	Increase elasticity of substitution between capital and labor in manufacturing.
E–3	Set minimum wage at 125 percent of unskilled wage rate.

TABLE 2—DISTRIBUTION OF HOUSEHOLD INCOME (PERCENT) AFTER TAXES AND TRANSFERS

Deciles (in percent)	Experiments														
	A–1	A–2	A–3	A–4	B–1	B–2	B–3	C–1	C–2	D–1	D–2	E–1	E–2	E–3	Base
Bottom 10	2.41	2.31	2.39	2.40	2.50	2.42	2.06	2.33	2.44	2.45	2.45	2.27	2.10	2.19	2.50
20	3.88	3.63	3.75	3.79	3.99	3.89	3.29	3.76	3.88	3.90	3.89	3.70	3.42	3.60	3.90
30	4.91	4.63	4.75	4.80	5.04	4.93	4.24	4.79	4.90	4.92	4.90	4.75	4.44	4.67	4.90
40	5.93	5.63	5.75	5.81	6.07	5.95	5.23	5.82	5.91	5.92	5.91	5.78	5.49	5.75	5.90
50	7.04	6.74	6.85	6.92	7.18	7.10	6.34	6.96	7.01	7.02	7.01	6.93	6.68	6.92	6.97
60	8.34	8.06	8.15	8.24	8.47	8.38	7.69	8.31	8.32	8.30	8.30	8.28	8.11	8.29	8.24
70	10.00	9.77	9.83	9.92	10.08	10.04	9.47	10.03	9.99	9.93	9.95	10.01	9.97	10.03	9.86
80	12.33	12.20	12.23	12.30	12.32	12.38	12.10	12.45	12.34	12.24	12.28	12.44	12.58	12.45	12.15
90	16.24	16.40	16.34	16.34	16.07	16.30	16.81	16.51	16.31	16.17	16.24	16.46	16.89	16.46	16.08
100	28.93	30.62	29.97	29.47	28.27	28.66	32.78	29.03	28.90	29.14	29.07	29.36	30.33	29.63	29.49
Gini Coefficient	.38	.40	.39	.39	.37	.38	.43	.39	.38	.38	.38	.39	.41	.40	.38

and, in particular, whether they are primarily urban or rural. There is also somewhat more scope for affecting the absolute level of poverty through economic intervention than for changing the overall distribution of incomes.

The remarkable insensitivity of the relative shares of national income accruing to households by deciles to the wide array of economic policy packages represented in the model suggests that, while one can make small gains in the welfare of the poor through large changes within the system, the goal of equity cannot be achieved without radical reform. This result is consistent with the Adelman-Morris findings.

C. A Major Conclusion

The two studies just described attack the issue of income distribution and de-

TABLE 3—PERCENT OF HOUSEHOLDS IN THE BOTTOM DECILE

Who are the Poor?

Socioeconomic Groups	Experiments														
	A–1	A–2	A–3	A–4	B–1	B–2	B–3	C–1	C–2	D–1	D–2	E–1	E–2	E–3	Base
Engineers	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Technicians	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	1.0	1.0	0.5	0.8
Skilled	14.2	3.9	1.5	6.5	3.7	9.5	2.2	17.7	14.4	7.6	6.9	9.0	16.8	3.9	8.8
Apprentices	0.4	0.2	7.6	0.3	0.2	0.5	0.2	0.0	0.5	0.3	0.3	0.5	0.6	0.2	0.3
Workers	4.3	12.4	7.9	13.8	13.5	17.0	5.4	15.2	16.8	19.1	15.1	5.6	28.6	10.5	19.4
White Collar	4.6	1.7	0.4	2.2	4.4	2.2	0.2	2.1	2.5	3.4	2.2	6.2	3.5	2.7	3.2
Government	8.6	6.4	3.0	9.4	11.1	13.0	8.0	11.0	10.3	9.2	11.8	6.9	4.6	6.5	7.4
Self-Employed 1	9.5	5.3	3.3	7.4	10.1	10.6	5.5	13.4	10.4	8.8	8.8	3.6	5.7	3.0	7.6
Self-Employed 2	11.1	7.1	3.1	9.7	12.3	15.9	7.7	18.7	14.2	9.8	12.7	5.4	4.9	5.2	9.8
Capitalists	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Agric. Labor	5.6	11.2	5.7	9.2	8.9	6.1	13.4	4.4	6.4	8.4	8.5	33.0	22.9	34.2	8.4
Agric. 1 (Tiny)	13.4	30.6	36.1	26.6	20.9	17.1	34.4	11.6	14.9	19.7	20.3	18.6	8.0	21.1	20.2
Agric. 2 (Small)	7.8	19.6	22.8	14.2	13.8	7.7	21.3	5.8	9.0	12.8	12.6	10.0	3.5	11.8	13.2
Agric. 3 (Medium)	0.4	1.4	8.5	0.7	0.9	0.3	1.8	0.2	0.4	0.8	0.7	0.3	0.0	0.3	0.9
Agric. 4 (Large)	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Average HH Income[a]	46.	47.	31.	51.	48.	52.	52.	52.	49.	47.	48.	46.	43.	42.	47.

[a] Average income is in thousands of real (1968) Won.

velopment with two very different techniques and bodies of data. The general conclusions they reach are very similar and therefore strongly reinforcing. Both studies show quite clearly that, in approaching the subject of economic development, whether theoretically or in terms of policy formulation, one cannot duck the issue of the purpose of development. This is a result of extreme importance: very different models, policy instruments, and disciplinary mixes are implicit in the choice of objectives. It is not simply a matter of tradeoffs; the optimal time sequences of dynamic strategies are almost certainly quite sensitive to the choice of goals. The conclusion is inescapable that, in setting the directions for the future of the field of economic development, we must be very clear, much clearer than we have been in the past, in our choice of development goals.

II. Development Objectives

The currently accepted definition of development focuses upon the creation of conditions for self-sustained growth in per capita *GNP* and the requisite modernization of economic, social, and political structures implicit in the achievement of this goal. To the extent that growth in *GNP* and concomitant changes, on the one hand, and improvements in the welfare of the poor, on the other, are monotonically and positively related (except possibly for short-term startup problems), this definition of development could offer an acceptable working definition. If, however, as is suggested by the work described above, a *U*-shaped relation exists, then the accepted definition of the development is, at the very least, badly misleading as a focus for thought and practice in the field. Since trickle-down does not work, basing the definition of the process of economic development upon a faulty analogy with modern industrial nations has led to the use of incorrect performance criteria, the selection of largely incorrect development strategies, and the use of inappropriately framed models and inappropriately focused planning methodologies.

On a moral basis, it seems to me that in each nation the proper long-term goal of national development policy must be the successive relaxation of the systemic obstacles to the full realization of the human potential of its members. The goals of *economic* development are then twofold: to provide the material basis for achieving these objectives and to establish the economic conditions for relaxing the other barriers to self-realization (access to education, work satisfaction, status, security, self-expression, and power). The proper balance between these two economic goals is a dynamic one and will depend on the state of welfare of the poorest segments of the population at each decision point.

This definition of the purpose of development, which I shall call ***depauperization***, is humanistic, largely spiritual, and highly dynamic. It is focused upon individual welfare, as perceived by the individual himself, with full recognition of the nonmaterial, human relations, and intergenerational aspects of personal welfare.

Depauperization has both economic and noneconomic dimensions and stresses the removal not only of material but equally importantly of social, political, and spiritual forms of deprivation. It involves not only equity but more significantly the creation of conditions conducive to continuing improvements in equity. It may require temporary sacrifices in national economic growth and involves major social, political and institutional change. In addition, the optimization of some of the components of this overall objective may, in the short run, necessitate sacrifices in some of the others, so that the ultimate realization of depauperization is likely to require appropriately phased sequences of dynamic strategies, each aimed at optimizing some subset of the total goal package.

In order to develop appropriate criteria, strategies, models, and methodologies for future development decisions, it is important to distinguish among four conceptual goals: equity, growth, equitable growth, and depauperization. Equity, growth, and depauperization are all special cases of equitable growth. Equity lies at one extreme. Equity alone as a goal, with growth given a subsidiary role, implies that the most important objectives are justice and fairness in some sense, rather than absolute welfare. Growth alone as a goal lies at the opposite extreme and has been discredited, as indicated above, in the minds of most of us. Depauperization, like equitable growth, involves some weighted combination of equity and growth, but in view of its explicit emphasis on the dynamic and nonmaterial aspects of individual welfare, depauperization lies well towards the side of equity. It implies that a primary objective of development is not only to help the poor to achieve a decent standard of living but in addition to create the conditions, economic and otherwise, that will allow them and their children continuing equal access to opportunities for self-realization. The achievement of depauperization, therefore, requires institutional change, a measure of economic, social, and political equity, and economic growth, not as goals in themselves but as instruments for the achievement of permanent self-perpetuating gains in the welfare of the poor.

A normative judgment implicit in depauperization is that no societal process can be morally acceptable which imposes, maintains, or intensifies the objective and perceptual barriers to self-realization of major segments of that society. To my mind, the major crime of humanity—practiced on an appalling scale—is to create and enforce, by means of institutional barriers and structural dynamics, conditions which lead to the waste of as much as 90 percent of its potential for fulfilled life, creativity, and personal growth. Empirical evidence shows that the natural historical process of modern economic growth works systematically to perpetuate those conditions for most of the period from the beginning of economic modernization. To counteract this social distortion should be a basic purpose of all national social policy and a prime objective of international assistance; to show how this might be done should be a major objective of all social scientists concerned with development.

III. Strategies for Depauperization

Most of the discussion concerning development strategies has been in terms of contemporary tradeoffs among major policy packages, rather than in terms of sequences of more or less pure strategies. A cursory survey of the dynamic processes that have led to the development of the currently industrialized nations suggests, however, that the development process has in fact proceeded as a sequence of pure strategies, rather than in a progression of mutually balanced strategy mixes. Emphasis on strategies that focus on a single major objective (such as the creation of a viable political system, or industrialization, or social development), largely unaccompanied by concomitant changes in the rest of the system, has in general created strategic imbalances which set conditions for the next phase. It is only at the beginning and the end of the entire process that there appears to be a balance among the various features of economic, social, and political development.

An examination of the development process of those non-Communist countries which have recently successfully combined improvements in the incomes of the poor with accelerated growth (Israel, Japan, South Korea, Singapore, and Taiwan) shows that they all have followed a similar dynamic sequence of strategies, varying only in the extent and consciousness with which it has been pursued. Aided by large infusions of foreign capital, all of these have followed the following dynamic sequence:

Stage I: Radical asset redistribution, focusing primarily on land;[2] but also imposing (at the very minimum) curbs upon the use and further accumulation of financial capital. This stage may involve negative growth rates, but is necessary to set the economic and political conditions to ensure that subsequent economic growth is not highly unequalizing.

Stage II: Massive accumulation of human capital, far in excess of current demand for skills. In this stage ownership of human capital is redistributed, the human resource base is vastly enlarged, and both the economic opportunities and the political pressures for the next stage are generated. In all five countries, this stage was accompanied by relatively slow rates of economic growth and, at later times, by political instability, social tension, and unrest.

Stage III: Rapid, human-resource-intensive growth. After the investment in human resources has been made, continuing depauperization requires that subsequent increases in growth rate be achieved through strategies that stress rapid labor-intensive growth. This implies that sufficient attention must be paid to the formulation of economic policy. In the smaller nations development will have to be oriented towards export markets. In large countries, on the other hand, industrialization can be oriented more towards satisfying domestic demand, particularly when a more equitable growth pattern generates a mass consumer market and when a more appropriate import-substitution technology is found.

To achieve depauperization, two extreme strategies are theoretically possible: grow now, redistribute and educate later; or redistribute and educate now, grow later.[3] The former strategy is the one that has typically been followed historically in developed non-Communist economies, except for the United States and Japan. The redistribute-first strategy is the one followed by the successful equitable-growth countries in the past two decades.

I am convinced that for rapid depauperization there must be a deliberate application of the redistribute-and-educate-first strategy on a wide scale. My argument is a practical one: there is selective but consistent evidence that this strategy can work and significant evidence that the opposite strategy does not work on anything resembling an acceptable time scale. Once the slower growth phase of Stages I and II is passed, the transition to a stage of relatively high development can be accomplished in short order. In the five countries considered, the first phase lasted about a decade. In another decade and a half, once economic incentives were restructured to permit an export-oriented industrialization that employed the new skills of the population, both depauperization and rapid economic growth became self-sustaining.

Some may dismiss the success stories as special cases, asserting that these are all small nations with nonrepresentative cultural traditions and attitudes, helped by unusually large per capita infusions of foreign aid and subjected to exceptional challenges that strengthened the legitimacy of their governments and made economic viability a major condition for national survival. Special cases they may be, but five successful cases are certainly more encouraging than none, and the consistency of their experiences surely weakens the "uniqueness" argument.

---

[2] Except in Singapore, which is entirely urban. A major welfare program, however, accomplished some of the same redistributive objectives.

[3] The remainder of this section is a minor reediting of Adelman (1974). While originally written as a prescription for equitable growth, I have since recognized that the prescription is more precisely one for depauperization. The recommendations for achieving a weaker form of equitable growth would differ primarily in the relative emphasis on growth and equity and the relative emphasis on social, political, and institutional changes compared with purely economic actions. The specific policies associated with achieving the respective goals would differ considerably more than the goals themselves.

There is also evidence that the entire package—resource redistribution, massive education, and labor-intensive growth policies—must be adopted in that sequence to achieve rapid success. Incomplete versions of this program, such as land reform alone or education without labor-intensive growth, have not worked. For the advanced countries which followed a grow-first pattern, economic development did eventually benefit the poor, but the time it took to do so was much longer (roughly two or more generations) than in our five successful cases.

The traditional strategy for economic development—rapid economic modernization involving trade and investment policies aimed at the fast expansion of the industrial sector—has not been so much wrong as premature for most developing countries. When applied *at the proper time in the proper strategic sequence* (i.e., *after* the equalizing preconditions of Stage I and Stage II), labor-intensive rapid industrialization phased in with appropriately paced agricultural development appears to be the key to realizing the full economic value of the redistributed assets and, hence, to *actual* depauperization, rather than to merely potential depauperization, as in Stages I and II.

REFERENCES

I. Adelman, "Strategies for Equitable Growth," *Challenge*, June 1974, 37–44.

——— and C. T. Morris, *Economic Growth and Social Equity in Developing Countries*, Stanford 1973.

——— and S. Robinson, "A Non-Linear, Dynamic, Micro-Economic Model of Korea: Factors Affecting the Distribution of Income in the Short Run," Woodrow Wilson School dis. pap. no. 36, July 1973.

G. Myrdal, *Value in Social Theory*, New York 1958.

The Structuralist Approach to Development Policy

By HOLLIS B. CHENERY*

Several approaches to the analysis of developing economies have evolved over the past twenty-five years. From a methodological standpoint, they can be grouped under three main headings: neoclassical, neo-Marxist, and structuralist. The first two attempt to adapt systems of thought that were initially formulated for the study of industrial societies to the less developed countries. The structuralist approach attempts to identify specific rigidities, lags, and other characteristics of the structure of developing economies that affect economic adjustments and the choice of development policy.

The initial set of structural hypotheses was formulated in the 1950's by writers such as Paul Rosenstein-Rodan, Ragnar Nurkse, W. Arthur Lewis, Paul Prebisch, Hans Singer, and Gunnar Myrdal. They explain phenomena such as balance of payments disequilibrium, unemployment, and worsening income distribution on the basis of particular properties of demand and production functions and other specifications of economic behavior. A common theme in most of this work is the failure of the equilibrating mechanisms of the price system to produce steady growth or a desirable distribution of income.

The success of a number of developing countries in accelerating their rates of growth in the 1960's casts some doubt on the significance of the structural problems that had been identified in the previous decade. However, in the past few years the importance of structural rigidities has been reemphasized by several new phenomena: the limited ability of economies to absorb the growing labor force, the worsening of the income distribution in several developing countries, and—most recently—the disruption to world trade caused by increased oil and food prices, which will require a substantial adjustment in productive structures. In short, development policy again seems to be constrained by a number of structural factors that require a more explicit analysis of the possibilities for short-term adjustment and for longer term changes in the economic structure itself.

In this brief survey, I will try to present an overview of structuralist methodology and to compare some of its policy conclusions to those of the other main approaches. Since there is not yet a unified theoretical framework for structural analysis, I will select formulations that seem to provide a promising basis for future research.

I. Methodology

The methodology of structural analysis has evolved over the past twenty-five years from a set of rather intuitive hypotheses to models of increasing empirical validity and analytical rigor. This evolution can be summarized in three stages: formulation of hypotheses, empirical testing, and the elaboration of more complete models. This sequence can be illustrated for two of the basic elements of structuralist systems: the concept of a dual economy and the concept of complementarity in demand, which underlies theories of balanced growth.

---

[*] Vice President for Development Policy, World Bank. I am indebted to Bela Balassa, Helen Hughes, Graham Pyatt, and Wouter Tims for helpful comments.

The concept of a dual economy stems from the observation that development takes place unevenly both within and between sectors of an economy. Although this concept has had many different formulations, the most influential is that of Lewis. He makes three basic assumptions as to the structure of a developing economy: (1) that technology can be divided between capital-using (capitalist) and noncapital-using (subsistence); (2) that the labor supply is elastic at a conventional wage; and (3) that saving is done largely by the recipients of nonwage income (capitalists). These assumptions, or variants of them, have been incorporated in models that explain the acceleration of growth, the allocation of the labor force, and changes in the distribution of income.

The early formulations of concepts of balanced growth by Nurkse and Rosenstein-Rodan also relied on a simple set of structural hypotheses: (1) a generalized version of Engels' law, specifying that consumer demand for food, clothing, shelter, and other major commodity groups is mainly a function of income and little affected by relative prices; (2) a similar assumption as to the limited price elasticity of demand for exports; and (3) in Rosenstein-Rodan's formulation, the importance of economies of scale in overhead facilities and basic industries. The first two assumptions make it necessary to expand output and allocate investment in close relation to the pattern of domestic demand. They also provide an explanation for structural disequilibrium and slow growth in countries that fail to do so.

Both sets of assumptions have in general stood up well to subsequent empirical tests. The acceleration of population growth has probably made the surplus labor assumption more generally valid today than when it was initially formulated for underdeveloped countries. Econometric tests by R. Weisskoff and C. Lluch and A. Powell provide some support for the generalized version of Engels' law underlying the theory of balanced growth, since they show most price elasticities to be less than unity. Among the basic structuralist assumptions listed, only the assumed inelastic demand for exports needs to be seriously qualified.

The third stage of theoretical refinement and policy application has proved more difficult. In the first place, it has been shown that the structural relations posited are not sufficient to lead to some of the policy conclusions suggested in the original formulations. As in the case of the Keynesian assumptions, a more complete formulation of models that can be statistically estimated has proved necessary in order to reach useful policy conclusions. Much current work consists in developing a second generation of models in the structuralist tradition that are designed for statistical application in individual countries, rather than for deriving broad generalizations.

The structuralist approach has had a substantial impact on both internal and external development policies. In both instances it focuses on identifying the consequences of various kinds of structural disequilibria. In domestic policy, the principal phenomena examined have been the effects of surplus labor on resource allocation and more recently the interpretation of worsening income distribution as resulting from a set of disequilibrium conditions. In international policy, analysis has focused on the nature of structural disequilibrium in the balance of payments and its effect on trade and aid policies. The following sections examine some of the relationships between the methodology employed and the policy conclusions reached in these two areas.

II. Internal Development Policy

Unlike the neoclassical assumptions, the structuralist alternatives do not lead automatically to policy conclusions. To produce such conclusions, they must also be embodied in an explicit general equilibrium framework. For this purpose, most analysts have used one of two simple models: either a neoclassical model with particular structural relations added or some version of a linear Leontief input-output model, which excludes most forms of substitution. The neoclassical framework minimizes the effects of the specific rigidities in the economic system, while the input-output system tends to exaggerate them.

Most elaborations of the dual economy and surplus labor concepts have been made in a two-sector neoclassical system, as in the work of J. C. Fei and G. Ranis, L. Lefeber, and A. C. Kelley, J. G. Williamson and R. J. Cheetham. One policy result from this type of model is to determine the shadow price of labor, which can then be used in project evaluation and in establishing the need for labor subsidies. Empirical applications of these models have been limited by the lack of data on different aspects of dualism. It will require some expansion of the two-sector framework before statistical estimation of the underlying structural relations becomes feasible.

The balanced growth hypotheses have been widely used in empirical analysis and are usually incorporated in input-output models which include foreign trade. These are applied to the formulation and testing of development plans, as illustrated in surveys by Chenery (1971) and A. Manne. Apart from its use in country planning, this type of model can also be used to deduce more general propositions through systematically varying the structural parameters and determining a set of solutions based on either simulation or optimization.

This form of sensitivity analysis lends considerable support to some of the conclusions of balanced growth theorists. For example, it is necessary in larger countries to expand agriculture at a rate that is largely determined by the income elasticity of domestic demand for foodstuffs because of the limited possibilities of expanding nonagricultural exports to offset a shortfall. Failure to meet this condition has retarded growth in a number of countries. This conclusion does not apply when industry is disaggregated to individual sectors such as steel or fertilizer, however, in which the optimal investment allocation to sectors having economies of scale is characterized by the alternating pattern of expansion of production and imports described by T. Scitovsky and by Chenery and L. Westphal.

The elaboration of structuralist hypotheses in planning models has also focused attention on the value of flexibility in adapting resource allocation to changing circumstances. This problem does not arise in the neoclassical system, which assumes perfect foresight and a high degree of substitutability. When these assumptions are abandoned, flexibility can be provided by increased exports or capital inflows and by planning some excess capacity in physical and human capital stocks. Although a formal treatment of the benefits of flexibility and the ways of achieving it has not been developed, this is an important problem for developing countries which cannot be analysed in the neoclassical system.

The revival of interest in income distribution has added a new dimension to structural analysis. The traditional approach focuses on the division between wage and nonwage income and is ill suited to developing countries, in which modern sector wage earners are in the middle-income groups. Recent studies by M. S. Ahluwalia have brought out the fact that the bulk of the poorest groups in developing countries are self-employed and largely rural. Their incomes depend more on the availability of land and capital and access to public facilities than on wages. Since each of the main poverty groups—small farmers, landless laborers, urban self-employed—has a different set of productive possibilities and constraints, a new form of structural analysis based on identification of these groups is necessary for distribution-oriented policies.

This recognition of the importance of asset distribution does not necessarily lead to the Marxist conclusion that the redistribution of existing assets is the only alternative. Measures to redistribute increments in income and new asset formation are more likely to be acceptable to the majority of the population and less disruptive of development in most countries. Development strategies based on this approach are elaborated by Ahluwalia and Chenery. The neo-Marxist policy recommendations suffer from the same defects as the neoclassical in that they are implicit in the initial assumptions rather than being derived from an analysis based on empirical estimates of the underlying structural relations.

III. External Policy

The conflict among the three analytical approaches is perhaps most acute in the area of external policy. The neoclassical approach tends to exaggerate the benefits of trade in an open economy when it does not explicitly consider the effects of uncertain export prices and the difficulties of shifting resources to meet changing market conditions. Conversely, the neo-Marxist approach exaggerates the costs of "dependence" on external trade and investment and tends to ignore the benefits of the technological transfers that accompany them. The early structuralist views have also proved to be excessively pessimistic as to the possibilities and benefits of nontraditional exports of manufactures and services.

The structuralist concept of development as characterized by rigidities that limit economic adjustments requires an analytical framework in which external policy is more closely linked to domestic resource allocation than does the neoclassical view, which minimizes these restrictions. Attempts to formalize these relationships started from simple two-gap models, which incorporate explicit limits on the rate of increase of domestic saving, investment, and exports. These models were elaborated by including optimizing procedures and shadow prices, the disaggregation of productive sectors, and more explicit treatment of structural change over time, as in the work of Chenery and A. MacEwan, M. Bruno, and S. D. Tendulkar. Experiments with these types of models in a number of countries have led to several general conclusions as to development analysis and policy, the most significant of which are: (1) the enhanced value of increased exports and capital inflows in bottleneck situations, in which the trade limitation is more restrictive than supplies of capital and skilled labor; (2) a restatement of the empirical basis for assessing comparative advantage over time in relation to the factors limiting development; and (3) clarification of the relation between internal and external constraints in the evaluation of individual investment projects.

During the 1960's a number of countries progressed from an initial strategy of import substitution to the promotion of manufactured exports after they had developed a sufficient industrial base to do so. As countries achieve a more diversified productive structure and reduce their concentration on a few exports, the difference between the neoclassical and structural prescriptions diminishes because some of the constraints that had previously limited growth are no longer significant. An assessment of the experience of some of the industrializing countries that have made this transition—Taiwan, Israel, Korea, Brazil, Mexico—suggests that earlier conclusions as to the high cost of the initial stage of import substitution need to be reexamined in light of the subsequent ability of the country to shift to manufactured exports and break out of the phase of trade limited development. This evaluation suggests that the dichotomy between inward-oriented and outward-oriented policies has perhaps been overdrawn, and that these policies can be more usefully viewed as sequential elements of a strategy designed to bring about changes in the structure of both production and trade.

The basic issues of the relations between trade and growth have been reopened in a new form by the substantial changes in relative prices that have taken place over the past two or three years. The main beneficiaries have been the oil-producing countries, which now have a structural problem of unprecedented magnitude in increasing their capacity to utilize their greatly increased export revenues. Most of the rest of the world has the opposite problem of adjusting its economic structure—through import substitution, increased exports, and redirecting trade—so as to accommodate a substantial worsening in its terms of trade. Not only has world income distribution been made worse by the present and prospective reduction in growth of the poorest countries, but within these countries the policies to improve income distribution are also being weakened. For most developing countries, increased import prices mean a return to a condition of a dominant trade gap and large capital inflows for the next few years. They need to give highest priority to reallocating resources so as to expand exports and gradually reduce capital inflows over a reasonable period.

This shock to the system of world trade has emphasized the advantages of the more flexible economies, such as Taiwan, Korea, and Brazil, which appear to be able to absorb large increases in import costs through a combination of increased exports and borrowing, while still maintaining substantial rates of growth. Conversely, the countries most severely affected by the rise in oil and food prices are those which are dependent on primary exports, whose prices kept up with world inflation for only a brief period, and which do not have the flexibility to shift readily to other exports. Unless oil prices are drastically reduced (which seems unlikely), substantial increases in capital flows to these countries will be needed over the next few years if they are to sustain even moderate rates of growth.

IV. Conclusions

The preceding discussion has illustrated the relations between theoretical premises and policy implications in several fields of development. The simplifying assumptions of the models currently in use tend to exaggerate the differences between neoclassical and structuralist prescriptions. As statistically determined relations replace a priori hypotheses, it is predictable that these differences will be reduced. A similar process can be hoped for when neo-Marxist theorists turn their attention to verifying their hypotheses.

Although better knowledge of elasticities of substitution in demand and production would do quite a bit to reduce the conflicts in policy guidance, several real differences in basic concepts remain. Neoclassical policy consists essentially in removing impediments to the functioning of markets so as to make the real world as much like the abstract model as possible. However, it will never be possible to achieve perfect knowledge or instantaneous adjustment to market signals. It is therefore necessary to incorporate these "imperfections" into the model itself. Once this has been done, it will become possible to take account of the existence of internal or external disequilibria and to devise more realistic policies to cope with them. In the theoretical literature, these policies are misleadingly referred to as "second best" in relation to the neoclassical model. It would be more accurate to characterize the model itself as overly simple and "first best" policies as simply unattainable. More attention should be given to improving the realism of basically neoclassical models instead of discarding them in favor of equally oversimplified structuralist formulations.

In minimizing or ignoring the advantages of market adjustments, structuralist (and Marxist) policy prescriptions usually put too much weight on the limited administrative apparatus of developing countries. As A. Hirschman has stressed, this is one of the main limitations to development; it should be allowed for by not seeking too much fine tuning of development policy. Difficulties in implementing a complex set of policies may prove much more costly than the allocative inefficiency of a simpler program that can be more readily carried out.

Finally, it must be recognized that the task of development has been made much more difficult for most countries by the recent changes in the world economy. Virtually every country is currently suffering from more or less serious disequilibria in its economic structure. Despite our desire to give greater weight in development policy to distributional considerations, this cannot be achieved without giving equal priority to adjustments in external trade and capital flows.

REFERENCES

M. S. Ahluwalia, "Income Inequality: Some Dimensions of the Problem," in H. Chenery et al., *Redistribution with Growth,* London 1974.

——— and H. Chenery, "A Model of Distribution and Growth," in H. Chenery et al., *Redistribution with Growth,* London 1974.

M. Bruno, "Optimal Patterns of Trade and Development," *Rev. Econ. Statist.,* Nov. 1967, *49,* 545–54.

H. Chenery, ed., *Studies in Development Planning,* Cambridge, Mass. 1971.

——— and A. MacEwan, "Optimal Patterns of Growth and Aid: The Case of Pakistan," in I. Adelman and E. Thorbecke, eds., *The Theory and Practice of Economic Development,* Baltimore 1966.

——— and L. Westphal, "Economies of Scale and Investment Over Time," in J. Margolis, ed., *Public Economics,* London 1969.

J. C. Fei and G. Ranis, *Development of the Labor Surplus Economy,* Homewood 1964.

A. Hirschman, *The Strategy of Economic Development,* New Haven 1958.

A. C. Kelley, J. G. Williamson and R. J. Cheetham, *Dualistic Economic Development,* Chicago 1972.

L. Lefeber, "Planning in a Surplus Labor Economy," *Amer. Econ. Rev.,* June 1968, *58,* 343–73.

W. A. Lewis, "Economic Development with Unlimited Supplies of Labor," *Manchester School,* May 1954, *22,* 132–91.

C. Lluch and A. Powell, *International Comparisons of Expenditure and Saving Patterns,* Int. Bank for Recon. and Dev., Res. Cent., dis. pap. no. 2, 1973.

A. Manne, "Multi-Sector Models for Development Planning: A Survey," *J. of Development Econ.,* June 1974, *1,* 43–70.

G. Myrdal, *Economic Theory and Underdeveloped Regions,* London 1957.

R. Nurkse, *Problems of Capital Formation in Underdeveloped Countries,* Oxford 1953.

R. Prebisch, "Commercial Policy in the Underdeveloped Countries," *Amer. Econ. Rev. Proc.,* May 1959, *49,* 251–73.

P. Rosenstein-Rodan, "Problems of Industrialization of Eastern and South-Eastern Europe," *Econ. J.,* June-Sept. 1943, *53,* 205–16.

T. Scitovsky, "Growth—Balanced or Unbalanced," in M. Abramovitz et al., *The Allocation of Economic Resources,* Stanford 1959.

H. W. Singer, "The Distribution of Gains Between Investing and Borrowing Countries," *Amer. Econ. Rev. Proc.*, May 1950, *40,* 473–85.

S. D. Tendulkar, "Interaction Between Domestic and Foreign Resources in Economic Growth: Some Experiments for India," in H. Chenery, ed., *Studies in Development Planning,* Cambridge, Mass. 1971.

R. Weisskoff, "Demand Elasticities for a Developing Economy: An International Comparison of Consumption Patterns," in H. Chenery, ed., *Studies in Development Planning,* Cambridge, Mass. 1971.

State of Development Economics

By Stephen A. Resnick*

The current state of development economics is reflected all too well by that of neoclassical theory—both are in a mess because of the lack of reality of the assumptions upon which they ultimately rest. What we now observe in the field of development is a focus upon reducing market imperfections, especially in the financial and trade sectors, and increasing investment in human resources. Development economists seem to have taken the institutional and historical development data into their academic offices, beaten the theory out of it, and come out with an *ad hoc* policy package of applied micro theory, no more useful in this field than in any other. Yet there has been a continuity to development theory that explains why it tends to be useless in explaining the reality of poverty and inequality and why, despite a passing appreciation of historical phenomena, it is a field that is totally ahistorical. What all of development economics since World War II have in common is the failure to see that the poverty of third world countries was never independent of the relative affluence of center countries. Specifically, the economies of the hinterland have been systematically organized and this organization historically reproduced by political and economic forces within the center to pump out unpaid surplus labor.

This alternative view is in complete contrast to what might be called neoclassical development theory for it analyzes unequal distribution of income as an essential and endogenous part of capitalist development. The major difference between the two approaches, about which there is much confusion, is the use of the concept "capitalism" to modify development and its derivative "imperialism" to analyze the process of international development.

The basic contrast between neoclassical and Marxian development theory must be found in the different underlying theories of value upon which each rests and from which each draws its logical framework and tools of analysis. In the neoclassical view of development, relationships of exchange not only dominate or determine production relationships, but take on a life of their own in explaining changes in a society over time. These exchange relationships depend upon preferences or choices of individuals as revealed in the market and ultimately upon some subjective concept of utility. This is perhaps a strong statement for rarely does one observe the explicit assertion that this is the theory being used. Nonetheless, upon what else is it to be based? Certainly an explanation based upon class analysis or the concept of exploitation would be an anathema to neoclassical development theory. The theory might be based upon some nebulous sociological notions of how primitive societies operate, but few economists would get caught holding this view. Perhaps all that really matters is the planning of development by the manipulation of carefully calculated input/output tables or the pragmatic use of incremental cost/benefit exercises. Even if one makes the heroic assumptions necessary to use the technical coefficients and somehow assumes that these coefficients are independent of government development policies, use of the technical

---

* Professor of economics, University of Massachusetts. I would like to thank Richard Wolff for his helpful comments.

planning approach, no matter how complex the models, must quickly come into conflict with the reality of national and international political relationships. Moreover, once the matrices are inverted, there remains the bothersome theoretical problem of what determines the wage rate of labor and/or the entrepreneur's decision to save and accumulate capital. Sometimes the power of technique lies in its ideological dimension which allows the economist as technician to ignore both the theory upon which it ultimately rests and the political economy that produces that theory.

What we currently observe is a careful analysis of those necessary reforms, taken from neoclassical theory of perfect competition, to allow the smooth workings of supply and demand in all sectors of the economy. The attempt is to resurrect and use the perfectly competitive model to explain not only why twenty-five years of import-substitution industrialization and growth of national output have not succeeded in remedying poverty, but also why, as has been commonly rediscovered once more, the distribution of income and wealth is getting more uneven over time. The view rests on the notion that the state through its financial and fiscal instruments has produced market imperfections that frustrate the development of free markets in both factor inputs and commodity outputs, and this has led to a less than optimal allocation of resources.

A policy package that deals with the more obvious imperfections, such as monetary and credit changes that allow the market rate of interest to reflect the opportunity cost of capital, foreign exchange devaluation and tariff changes and major reductions in quantitative restrictions on foreign trade that produce a more uniform ad valorem tariff structure directed toward reducing overall levels of protection, and tax and expenditure reforms that remove differential incentives to favored groups, will ultimately allow a situation where market signals perform their proper allocative role. Once accomplished, the distribution of income will be determined by the logic of the impersonal market, such that each factor receives what it deserves in terms of its marginal contribution to output.

These changes are seen as gradually moving the economy to its potential production possibility frontier by allowing it to find its Ricardian comparative advantage in the export of labor-intensive manufactures and as facilitating the growth of agricultural output by removing distortions that discriminate against it. Reforms are also seen as releasing a good deal of entrepreneurial talent, especially that bottled-up in the so-called informal urban and rural sectors.

Neoclassical development has moved to this emphasis on the freeing of markets to reflect opportunity costs and consumer preferences from a previous approach that placed more emphasis on institutional change per se, big pushes whether in agriculture or industry, and the breaking of vicious circles. Yet the differences are more apparent than real for both views share the common faith that the developmental effort must be to minimize those barriers to the effective operation of supply and demand, whether caused by cultural factors, indivisibilities, or by government-induced distortions, so as to maximize the individual private enterprise of each agent in the society. Thus, these views are quite consistent with neoclassical theory.

What is wrong with this approach? First, the theoretical argument is invariably confined to the realm of exchange relations. It omits the underlying national or international relations of production out of which emerge a specific institutional environment and a specific set of supply and demand equations. Second, the policies advocated, consistent with the theory, are seen as necessary reforms of distorted internal market relations that will allow the economy to independently break out of the postwar industrialization experience to compete effectively on the international market. However, this very international market serves as the principal vehicle by which units of production in the center exploit those in the hinterland and prevent their independent development.

Simply put, the neoclassical approach is misspecified because of the omission of production relations and thus yields biased policy conclusions and unreliable predictions. Further, although this approach has recently appended to its analysis the more obvious social and political issues, they are added as unexamined external givens never seen as the direct outgrowth of the underlying structure of production, i.e., the value relation between labor and labor power. Neoclassical development cannot analyze anything outside of a framework of market or exchange relationships because that is the theory upon which it is based; it is trapped not by inadequate data or lack of "better" models, but rather by its narrow focus on supply and demand and its total neglect of those historic forces that have produced international relations of production and technology based upon an exploitive system of one class over another.

In contrast, a Marxian approach to development begins with a theory of value based upon relations of production determined historically; it is these relations that determine or govern those of exchange. What we observe in the hinterland are those dependent social relations of production and underdeveloped forces of production produced historically and reproduced currently by international capitalist development. Capitalism develops the hinterland not by underdeveloping it in the sense of making producers independent of markets, but rather by integrating it into the world market. It is the specific role of imperialism to systematically develop, integrate, and coordinate production in the hinterland into a specific pattern of international commodity and capital flows. Historically, imperialism produced a complex network of production and finance that continually operates so as to keep the hinterland dependent on the center. International trade becomes the economic mechanism that links and binds together modes of production in different areas of the world, and the actual or potential use of force is the final political weapon that prevents them from falling apart. Thus it is the capitalist mode of developing the international division of labor that makes commodity production a universal category, thereby creating one capitalist world comprising two unequal but dependent subsections. In this process, the hinterland's needs and tastes are divorced from its own history and its own social development and, along with comparative advantage and prices, are made a function of the center's economic growth and political power.

Relationships between center and hinterland, however, are not static, but must be understood within the changing process of capitalist development. Otherwise what seems to be exogenous to the hinterland is in reality the internal development of international capitalism. Specifically, the quantitative and qualitative changes that have occurred since World War II in the hinterland cannot be understood or analyzed properly independently of the historic process of capital accumulation in the center.

Although imperialism has a long history, the form of exploitation has changed. In the early phases, the merchant capitalist often used coercion directly to pump out surplus labor from peasants. Plunder, looting, and slavery were used in this process of slash and burn capitalism to create those fortunes that helped to finance the primary accumulation of capital in the center. Even during this period, there began to develop the exchange of food and raw materials from the hinterland for manufactures produced in the center's growing factories. The laws of commodity exchange started, but became the dominant characteristic between center and hinterland only when capitalist development emerged into its next phase of absolute concentration of production. From about the last third of the nineteenth century to World War II, the long-run tendency for the profit rate to fall in the center, as a result of increased competition among capitals produced by the accumulation and centralization of capital, became a basic contradiction of the system and led to a new phase of imperialism. It took the form of harnessing and then changing a precapitalist mode in the hinterland to the new raw material and export requirements of the developing capitalist mode in the center. International trade and finance became important mechanisms in combatting the tendency of the profit rate to fall.

Colonialism established the necessary social and physical infrastructure for the primary accumulation of capital within the hinterland. A variety of political and economic instruments were used to create an available and cheap labor supply for export production and to widen the market for the center's output. Thus, the colonial state produced those internal forces that led to the effective organization of supply or cost conditions, and the demand for commodities was governed by external forces dependent on the accumulation of capital in the center. What a colonial country imported and exported, and at what price, or what emerged as its comparative advantage, was dependent upon political and economic forces within the center over which it had no control.

There is an important change in the third phase of imperialism corresponding to the transition from absolute concentration to monopoly capitalism in the center. The interest of capital in combatting the tendency of the profit rate to decline has not abated, but its international aspect has taken on a new form. Formal colonialism no longer exists. The multinational corporation, itself a product of this transition, has become the outstanding feature of the new international system that continues the historic dependence of the hinterland upon the center.

National independence after World War II produced a situation where the state was first controlled and then used by a precarious coalition of landed and merchant interests, in whose hands wealth was historically concentrated during the colonial process, to create a profitable environment for the transfer of their nonindustrial assets into industrial claims on wealth. The state through a variety of trade and fiscal instruments endeavored to manufacture a class of entrepreneurs who would be willing to hold and accumulate industrial assets.

Differential access to credit, foreign exchange earnings, import licenses, and tax incentives for a favored class indicates the state's endeavor to develop industry within a ruling oligarchy of landed and merchant wealth. Market imperfections indeed existed, but were created for the use of one class over another. The rate of return for this class was high, as indicated by the growing income gap, so long as an alliance between rural and urban poor was prevented. Thus, repressive governments are required to keep the rural and the mass of nonprivileged urban workers in place while an internal colonial policy is used to foster the accumulation of capital.

This industrialization effort has not produced an independent capitalistic class nor has it made the economy independent of the international economic system. In fact, during what might be called round one of the postwar period, the very type of forced industrialization experienced had deepened dependence on the center rather than reduced it. Increased imports of raw materials and capital goods and reliance upon the technology of the center are the results of import substitution. Indeed, the very survival of much of the manufacturing sector depends upon foreign inputs and, thus, the expansion of exports to finance these needs.

The national bourgeois class is as dependent upon foreign capital and its political power as is the industry established. The second phase of imperialism acted as a barrier to the development of local industrial capitalism and produced instead a class of traders, landed aristocrats, and government officials who had all the trappings and mentality of merchants, individuals more interested in the maintenance of their privileged position in the sphere of exchange than in seeking new forms of production to invest their accumulated surplus. They were the capitalistic children of the bourgeoisie of the center and would require a protected environment to develop, but they could not and cannot outgrow the environment and create one of their own. Their industrial experience begins within a setting of state-fostered oligopoly and monopoly power rather than competition, perhaps partly due to technical reasons of optimum plant size, but more likely because of the need for the state, as an engine of growth, to overcome a disadvantaged past. The process of capitalist development followed produced a new era of industry, but reproduced their dependence upon international relations of production over which they have no basic control.

The multinational corporation has become the chief agency in this third phase of imperialism, incorporating the local bourgeois class into its corporate hierarchy, as at best junior partners, and linking and integrating the hinterland's industry into a global network of capitalist production. Whatever legal organization is used to conform with national laws of ownership, capitals of the center use a variety of institutional mechanisms, e.g., foreign subsidiaries, joint ventures, and corporate devices (management fees, patent and franchise royalties, intracompany pricing, marketing fees) to effectively transfer profits to the center and maintain control over the production process. National industry may appear to be growing, but it is little more than national plants of multinational corporations whose local "owners" may have control over the day-to-day operations, but not over decisions on pricing, output, and investment. The structure of Canadian industry may be the future for the hinterland, but at present it looks more like Puerto Rico, a branch plant country of United States capital. In either case, the hegemony of the center has not been diminished. Yet for some countries, round one may be ending.

Taiwan and South Korea may be forerunners of a new form of dependent capitalist development in round two. The organization of industry, dominated by foreign capital, has already been established, and cheap labor power has been made available for wage labor employment in both agriculture and industry through the fragmentation of the natural economy in the agrarian sector. Politically these countries represent safe havens for capitals of the center to exploit the existing cheap labor supply because of their dependence on *U.S.* military assistance. This political stability and the economic reforms passed make it attractive and profitable for multinationals to exploit lower unit labor cost in these countries, as compared to the center, to produce and export manufactured commodities to the center and to other countries of the hinterland. Taiwan and

South Korea may represent a preview of the future as the import-substitution phase of round one gives way to the exports of manufactures of round two. This process completes both the dependency of production in the hinterland upon that of the center and the reproduction of colonialism without formal colonies.

The observed rise in employment of wage labor in these countries, generated by the rise in exports, indicates the qualitative and quantitative change in social relations of production. Individuals are increasingly forced to have the freedom to sell their labor power in the market. Correspondingly they become subject to the domination and uncertainty of international capital. The drain of surplus value continues. No doubt this new era is already undergoing change as frictions appear among competing capitals of the center, and the very proletarianization of the hinterland's labor force produces an endogenous agent of change. More analysis is needed to understand these forces now emerging. Yet the very lack of neoclassical development theory to understand how one class exploits another, duplicated in its lack of understanding of how one country can exploit another, makes it useless in this new endeavor.

WHAT HAVE WE LEARNED FROM ECONOMIC GROWTH THEORY?

Implications of Growth Theory for Trade and Development

By RONALD FINDLAY*

This brief survey of the implications of growth theory for trade and development is in three parts. The first discusses the Ricardian and Marxian approaches, the second is devoted to some well-known contributions of modern growth theory, and the third offers a hypothesis as to why all three frameworks are each inadequate to serve as explanatory schema for the major trends in the relations between the "center" and the "periphery" of the world economy.

I.

In assessing the impact of growth theory on trade and development, I find it useful to consider the doctrinal history of the subject. Modern growth theory is usually taken to begin with R. F. Harrod's essay of 1939, but we should not forget that David Ricardo was a growth theorist too. Technical relations between inputs and outputs, propensities to consume and to save out of different classes of income, and the determinants of the growth of the labor force over time were all fairly explicitly specified and the implications for the growth and distribution of the national income were rigorously deduced. While the assumptions chosen were of course radically different from those of contemporary writers, the analytical procedure and method were the same in spirit and in form.

The basic Ricardian model is usually formulated as a sequence of harvest cycles in an agricultural economy in which output is a function of the supply of land which is fixed and the supply of labor which is proportional to the stocks of corn held by capitalists as "advances to labor." Landlords get the marginal product of land for each acre that they rent out, and capital gets the marginal product of the labor that it supports minus the subsistence wage. Rent is all consumed and profits are all accumulated so that the labor force expands, raising the marginal product of land and hence total rent, while reducing its own marginal product, and hence the margin of profit for the capitalists, until it falls to the subsistence wage itself, at which point profits become zero and the stationary state is reached. The accumulation of circulating capital, which drives the system forward, is eventually checked by the inert force of something that cannot be indefinitely accumulated, the "original and indestructible powers of the soil." The Ricardian capitalist to whom, as Marx was to say later, accumulation was "Moses and the prophets" is ultimately forced to the ultimate indignity, which is to get back at the end of the year merely what he advanced at the beginning.

The way in which international trade can modify these long-run implications of the basic model is analyzed by Ricardo in his *Essay on the Influence of a Low Price of*

---

* Columbia University.

Corn Upon the Profits of Stock. His discussion there is of much greater interest in connection with the growth-trade nexus than his much more famous static analysis in the *Principles* involving the celebrated numerical example of England and Portugal exchanging cloth and wine. In this analysis, as formalized in Findlay (1974) drawing on the work of L. J. Pasinetti, there are two sectors in the economy, agriculture, with the characteristics described above, as well as manufacturing, which produces luxury consumption goods purchased by the landlords out of rent. Relative prices are determined by the requirement that the rate of profit be equalized between the two sectors. Repeal of the Corn Laws or opening the economy to international trade would result in less labor being used in agriculture, lowering rents, and raising the rate of profit. Manufacturing output will expand to provide the exports necessary to pay for the corn imports. Thus it would seem that international trade can avert the approach of the stationary state, since the higher rate of profit would permit a higher rate of accumulation. However, land is limited in the long run in the trading partners as well and so world accumulation and labor force growth must raise the relative price of corn in comparison with manufactures, thus raising rents and lowering the rate of profit. Ultimately the world economy as a whole must tend towards a stationary state. Trade can, therefore, postpone but not abolish the cessation of growth that results from diminishing returns.

It is interesting to observe that the opening of trade influences the rate of profit and hence the rate of growth in the opposite way for the two partners, raising it for the exporter of manufactures and lowering it for the exporter of corn, the reason being that the shift in income distribution is favorable for accumulation in one case and unfavorable in the other. The static gains from trade are shared mutually, but the dynamic effects are inversely related. The terms of trade, however, move secularly in favor of the agricultural exporter.

In spite of its obvious shortcomings, one must admire the way in which Ricardo was able to provide a simple analytical framework to relate growth, income distribution, and trade in such a sweeping way. His was truly a "magnificent dynamics."

Growth theory in the grand manner was continued by Karl Marx. In spite of the many Ricardian aspects of his thought, one fundamental difference between their systems is the fact that Marx did not bring the limited supply of land and diminishing returns into his scheme in any essential way. His biographers have informed us of Marx's love of Greek tragedy and perhaps his unconscious dramatic instincts rebelled at bringing down the tragic hero of his system, the capitalist entrepreneur whom he both admired and hated, by something as crude and obvious as limits of cultivable land. Like all good German romantics, Marx thought that nature could be transformed and could not by itself constrict the limitless progress of humanity. Technical progress, the genie in the bottle that Ricardo never found, comes on to the center of the stage in Marx's vision of capitalist development. Like the tragic hero, it is not external circumstances of environment that would destroy capitalism but the fatal flaw within, the "internal contradiction," as he would say. Capital accumulation and technical progress expand production possibilities while the "immiserization of the proletariat" that the process is supposed to create makes it increasingly difficult for the additional output to be marketed, leading to cyclical crises and depressions that get increasingly severe until they finally produce a proletarian revolution.

The international aspects of capitalist development and its impact on traditional societies produce ambivalent reactions in Marx. On the one hand he condemns the plunder and disruption that it brings, but he also welcomes the "battering down of Chinese walls" and the end of the millennial stagnation of the so-called "Asiatic mode of production" in which a superstructure of despotism and bureaucracy rests on a basis of self-sufficient village production, maintained by forced labor on irrigation and other public works. The British destroyed the weavers of Bengal, but they also brought the railways, and in Marx's opinion, India would never be the same again.

Later Marxist writers, including of course Lenin, brought in foreign investment and associated export booms as a way in which capitalism attempts to overcome its internal contradictions of insufficiency of effective demand due to the impoverishment of the proletariat. This literature stresses the negative aspects of the impact of capitalism on less developed areas, in which imperialism is seen as suppressing incipient industrialization in the interest of maintaining markets for the colonial powers. The most recent major contribution in this tradition was the stimulating work on the multinational corporation of Stephen Hymer, whose tragic death robbed our profession of one of its all too few original points of view.

II.

Growth theory was never a systematic concern of the major neoclassical writers and its modern revival by Harrod and E. D. Domar was inspired by Keynesian considerations of how to maintain adequate effective demand to match a growing productive capacity and labor force. Domar's contribution was to show that investment had to grow exponentially at a rate equal to the savings rate divided by the capital-output ratio if it was to generate the effective demand required to fully utilize the productive capacity that it created. Harrod contributed the concept of the "natural rate of growth" equal to the rate of population increase plus the growth of productivity, both assumed exogenously fixed, to which the actual growth of the economy had to conform if instability was to be avoided.

Though Harrod and Domar each came to it in different and subtle ways, their names, if not their ideas, were merged into the simple formula that the rate of growth of output was equal to the savings rate divided by the capital-output ratio. This formula was widely used in the various "development plans" that proliferated in the 1950's and 1960's. Typically a target growth of per capita income was added to the estimated rate of population growth to get a desired growth rate, which was then divided by an assumed capital-output ratio to get the required rate of saving. Fiscal and foreign assistance measures were then set to achieve this required savings ratio. Though it is now somewhat fashionable to deride these efforts, I believe that the widespread acquaintance with this simple relationship on the part of planners and administrators contributed towards coherence and consistency in the aggregative framework in which development planning was conducted. The reasons for failure or relative lack of success in most cases was inability to predict sufficiently accurately such things as the demand and supply of foreign exchange and the absence of adequate machinery for implementation and control at the project level.

Using the simple formula as it stands would seem to imply that growth is proportional to the savings rate and the reciprocal of the capital-output ratio. The constraint of the natural rate of growth, however, could be argued as not binding in economies with extensive rural and even urban underemployment. In a celebrated paper W. Arthur Lewis took an explicitly Ricardian approach to the problem of growth in a developing economy, arguing that the real wage of the "modern" sector is set by conditions in the much more heavily populated "traditional" sector of a "dual" economy. The growth rate of the modern sector then depends upon the fraction of profits that are saved, assuming that the wage is too low to permit any saving by workers. Employment can expand at a more or less constant wage rate in step with the growth of capital and at a rate above the natural rate of growth until such time as the underemployed pool of labor or "industrial reserve army" has been depleted.

The relationship between the growth rates of capital and output and the natural rate in a fully employed economy was determined by R. M. Solow (1956) who used the concept of a constant returns to scale aggregate production function with capital and labor as the inputs. He showed that the Harrod-Domar formula determined the growth rate of capital, with the growth rate of output a weighted average of the capital growth rate and the natural rate. With the capital-output ratio free to vary in response to changing relative factor prices, the capital growth rate converged in the long run to the natural rate. Hence, increases in the rate of saving could not permanently raise the growth rate of output above the natural rate, though the level of per capita output would be higher in the long run the higher the constant rate of saving. This result, together with the calculation in Solow (1957) showing that apparently 85 percent of the growth in per capita output in the United States was due to technical change instead of capital deepening, had the implication for development policy that the main thrust over the long haul would have to come from improvements in technology.

Since the developing countries presumably are able to draw on a vast backlog of new technology already available in the more advanced countries, this view could be regarded as having optimistic consequences for the growth prospects of the former group. The "green revolution" excitement of the early and middle 1960's reflected this enthusiasm about the possibilities of technological change for development. However, it soon became clear that the diffusion of technological change is itself a complex and resource-using process, and moreover, one about which we know very little. The quality of the labor force in the receiving country is one obvious factor influencing the rate at which technology can be borrowed, making a case for the priority of investment in human instead of physical capital. Much has also been made of the multinational corporation as a vehicle for the transfer of technology. Very little analytical work has been done in this area. It would seem that one possible approach is along the lines of a paper by E. S. Phelps and R. R. Nelson. The rate of technological change in the borrowing country could be made a function of the gap between the current levels of its own technology and that in the advanced countries, which can be assumed to change at some exogenous rate, the rate at which borrowing takes place depending upon such factors as the educational level of the labor force and the extent to which plants run by multinational corporations operate within the domestic economy, providing an opportunity for "learning by seeing."

The two-sector version of the neoclassical growth model presents a convenient framework for the formal integration of growth and trade theory. The Heckscher-Ohlin factor endowment approach to international trade was given its familiar two-factor, two-good specification in the 1930's and 1940's by Abba Lerner, Paul Samuelson, and others. H. Oniki and

H. Uzawa made one sector the capital goods sector and the other the consumption goods sector and produced a model of trade and growth in a two-country world economy with a common natural rate of growth for the two countries. The pattern of comparative costs changes over time in such a model along with the relative capital-labor ratios. Comparative advantage is thus not determined once and for all by a given endowment as in the traditional approach, but instead depends upon parameters such as the rate of saving and the growth rates of population and technological progress that can, within limits, be altered by policy.

The exclusion of natural resources from this model seriously restricts its applicability and relevance to such problems as the secular tendency of the terms of trade. As we have seen, the Ricardian model predicts that this will be in favor of the primary exporter while writers such as Raul Prebisch have taken the view that technological progress in economizing on raw material requirements and the substitution of synthetic materials for natural products more than offsets the effect of diminishing returns in the opposite direction. The neoclassical growth models as they stand are not able to even consider this important question.

For further investigations of trade and growth in the context of neoclassical and dualistic models, the reader may consult P. K. Bardhan and Findlay (1973).

III.

Growth theory, like the theory of anything else, has to be confronted with the relevant facts. In the present case the facts are the economic history of the last two hundred years or so since the Industrial Revolution in England transformed the world. What precise combination of economic, social, and other forces brought that singular event about is impossible to ascertain in any precise way and is beyond the scope of a narrowly technical subject like growth theory. Marxists may claim to have an answer, but one searches the literature on the "transition from feudalism to capitalism" in vain for any endogenous mechanism to account for it. Alternative theories of Max Weber and others are no more satisfying. Given that one exogenous impulse, however, one would expect to be able to construct a theoretical scheme that will do some justice to the manner in which the frontier of "modern economic growth," as Simon Kuznets calls it, spreads from the initiating "center" of the world economy to a receding "periphery" until the globe as a whole becomes a single interdependent system. Can growth theory be said to have provided, even modestly, the explanatory power that one would expect of such a scheme?

The answer, I think, must be in the negative and it is interesting to speculate why. One important reason, it seems to me, is that it is impossible to make any sense of this process without considering it in terms of a complex interplay between technological progress and the availability of natural resources. Yet, strange as it may seem, none of the three broad approaches to growth theory that we have surveyed looks at the problem in terms of *both* these fundamental determinants. Ricardo, we tell our students, was led to false prognostications because he ignored technical progress. Yet contemporary growth theorists, of either Cambridge, blandly ignore any mention of land or raw materials, assuming that somehow technological progress can ensure that steady growth can go on forever. To her credit, Joan Robinson refers to such a state of affairs as a "golden age" to remind us of its mythical character, but the tough problems of how to deal with natural resources in models of growth do not seem to have been faced by anybody. The current crises through which we are passing and the profound alteration in the balance of world economic power that is taking place are all manifestations of the simple fact that the earth's resources are finite, something that growth theory since Ricardo never seems to have recognized. David Landes suggestively entitled his fascinating book on the history of technological change since 1750 *The Unbound Prometheus*. We should not forget the price that Prometheus had to pay for stealing the fire from heaven.

REFERENCES

P. K. Bardhan, *Economic Growth, Development and Foreign Trade*, New York 1971.

E. D. Domar, "Capital Expansion, Rate of Growth and Employment," *Econometrica*, 1946.

R. Findlay, *Capital Theory and Development Planning*, New York 1973.

———, "Relative Prices, Growth and Trade in a Simple Ricardian System," *Economica*, Feb. 1974.

R. F. Harrod, "An Essay in Dynamic Theory," *Econ. J.*, 1939.

D. Landes, *The Unbound Prometheus*, London 1969.

W. A. Lewis, "Economic Development with Unlimited Supplies of Labour," *Manchester School*, May 1954, *22*, 132–91.

H. Onki and H. Uzawa, "Patterns of Trade and Investment in a Dynamic Model of International Trade," *Rev. of Econ. Stud.*, Jan. 1965.

L. J. Pasinetti, "A Mathematical Reformulation of the Ricardian System," *Rev. of Econ. Stud.*, Feb. 1960.

E. S. Phelps and R. R. Nelson, "Investment in Humans, Technological Diffusion and Economic Growth," *Amer. Econ. Rev. Proc.*, May 1966.

D. Ricardo, *Essay on the Influence of a Low Price of Corn Upon the Profits of Stock*, London 1815.

R. M. Solow, "A Contribution to the Theory of Economic Growth," *Quart. J. of Econ.*, Feb. 1956.

———, "Technical Change and the Aggregate Production Function," *Rev. of Econ. and Statist.*, Aug. 1957.

The Theory of Economic Growth: A Critique and Reformulation

By DONALD J. HARRIS*

What has become of the modern theory of economic growth? In its dominant neoclassical form, the basic idea was that a capitalist economy is inherently capable of undergoing a sustained equilibrium of steady growth with divergences that are "fairly small, casual, and hardly self-accentuating" (R. Solow). This idea was at first starkly portrayed in a well-known parable constructed expressly for the purpose of conveying it. The idea was supposed to remain essentially the same after all the complications excluded from the simple parable were introduced. It turns out, however, upon further analysis, that this presumption is unfounded, once allowance is made for the heterogeneity of means of production which characterizes an advanced division of labor (E. Burmeister and A. R. Dobell, chs. 8, 9). Under these general conditions, not only is the neoclassical growth equilibrium not necessarily unique, but also the stability of the approach to any such equilibrium cannot be guaranteed. Efforts to resolve these problems have so far involved a resort to *ad hoc* assumptions that are unsupported by any systematic theoretical justification.

Admittedly, these analytical problems which the neoclassical theory faces are quite serious at a formal level. Moreover, they cannot be escaped by abandoning the simple parable as a bastard child and retreating into the supposedly more legitimate realm of neoclassical general equilibrium theory. For if there is such a thing as a neoclassical theory of economic growth, the substance of it must be capable of some form of coherent expression. Otherwise it remains the hollow shell of a system of differential equations.

Nevertheless, viewed on the plane of the system of neoclassical theory as a whole, these problems are of a relatively minor nature. On this plane the crucial question is, rather, whether that system of theory, *qua* theory, has anything at all to say about the process of accumulation in a capitalist economy. In its original and basic conception, neoclassical theory constitutes the process of accumulation as a process of "deepening the structure of capital," in which the saving decisions of atomistic individuals drive the economy to a stationary state where the ratio of the quantity of capital to labor is optimal and where there is no further inducement to save because the rate of profit (or rate of interest) has fallen to the level of the rate of intertemporal preference or the "cost of waiting" (K. Wicksell, J. Hirshleifer). What drives the economy in this conception are the intertemporal preferences of individuals engaged in exchanging consumption today for consumption tomorrow. The capitalist firm is seen merely as an intermediary between the individuals as suppliers of factors and the individuals as *rentiers* consuming their lifetime income. The basic objective of the economy as conceived in such a system is consumption; accumulation is an incidental feature of individuals' consumption decisions and comes to an end in the stationary state. The theory in this way denies the self-sustaining drive for expansion of capital as an essential feature of the capitalist mode of production.

---

* Associate professor, University of Wisconsin.

Continued expansion occurs in the neoclassical system only if it is assumed that there is an autonomous increase in factor supplies, consisting of the labor force, and autonomous changes in technology, both of which are unexplained *data*. In this respect, if continued expansion occurs it is for reasons that cannot be found within the theory and hence there is no *theory* of growth at all. Furthermore, the process of expansion (it is not so much a process as it is a "sequence of momentary equilibria"), to the extent that it occurs at all, is conceived essentially as a matter of piling up stocks of means of production (specific use values) either in the form of "putty" or as a heterogeneous collection of physical objects under the reign of Say's Law in an economy of simple commodity production (Burmeister and Dobell). In such a world, crises of overproduction and unemployment, booms and depressions, which are recurrent features of the entire history of capitalist development, do not arise except as temporary aberrations of the system. There is no identifiable class of workers displaced from property in the means of production who must depend entirely on employment in capitalist production, with its ups and downs, for their economic survival. It is therefore difficult to see what real historical phenomena, if any, this system of thought is intended to explain.[1] The central analytical question which is treated throughout all variants of the theory remains, not one of the historical laws of motion of capitalism, but rather one of the efficient allocation of resources in accordance with consumers' preferences (E. Malinvaud). This is evidently the whole basis of the singular concern of neoclassical theory with the problem of optimal growth.

It is not possible to pursue further in this short space the direct critique of the neoclassical theory. For this it would be necessary to locate in the basic conceptual structure of that theory, in the neoclassical theory of subjective value, the internal obstacles to its ability to grasp the process of capitalist production as a whole. This work is being pursued in different ways elsewhere (Harris, 1973 and 1974) and by others (e.g., J. Robinson, 1971, D. Levine, and E. K. Hunt and J. G. Schwartz). The time has come, however, to seek to establish the theory of economic growth on an entirely different footing. Some partial elements for such a reconstruction of theory are to be found in recent work, especially that of Robinson, and in earlier work of M. Kalecki and J. Steindl. But for a truly fundamental and integrated system of analysis, we have to turn to the tradition of Karl Marx who was the theorist of economic growth *par excellence*.[2]

Speaking broadly and briefly, a theory of economic growth is to be conceived as an explanation of the causes of the contradictory development of the capitalist mode of production, based on the observable historical reality of an immense expansion of productive forces and revolution in methods of production under capitalism combined with the persistence of unevenly developed sectors over large areas of the capitalist world and with periodically recurring crises affecting all or most of the major branches of industry. What is to be explained is this specific form of capitalist development: the nature of the forces which propel the system forward in this way and which account for the particular (contradictory) form of its forward movement. A theory of growth is, in this view, *a theory of the expanded reproduction of the capitalist mode of production on a world scale*. There are three central and interrelated elements involved in the construction of such a theory. These are: (1) the process of production of value and surplus value, at the heart of which is the conflicting *social* relation of capital to labor beginning in production and extending to other spheres; (2) the process of capital accumulation involving a continual drive for expansion of capital accompanied by changes in methods of production; and (3) the role of the state as dictated by the requirements of reproduction of capital and the contradictions associated therewith.

---

[1] If we examine as against this the humble beginnings of the modern theory in the works of R. F. Harrod, E. D. Domar, Sir John Hicks, and others of the business cycle theorists (see R. A. Gordon and L. R. Klein), we find that this work, though it did not get very far as a whole, was genuinely concerned to explain real phenomena in contemporary capitalism. Among other things, it failed specifically to integrate a theory of growth and the business cycle as aspects of the same phenomenon. In later work the problem of the business cycle was altogether displaced by a preoccupation with the properties of steady-state growth. There is no mention of that problem at all in N. Kaldor's definitive list of the "stylized facts" appropriate to the construction of a theory of economic growth.

[2] I exclude here the tradition of classical political economy, despite the fashion of holding up that tradition as the paragon of growth theory. This is because the classical economists failed to consummate a theory appropriate to the conditions of industrial capitalism. Their ideas were essentially limited to the conditions of an agrarian capitalism, without significant change in methods of production (or foreign trade), in which, due to the limited quantity and diminishing fertility of the soil, growth is arrested by increasing costs of production of wage goods. The Malthusian variant of this theme has, rather curiously, risen to prominence again in recent times.

There is no room in this view for conceiving of a separate and distinct theory of underdevelopment, as commonly understood, except as a theory of the transition from noncapitalist production to capitalism. Some glimpses of the latter are beginning to emerge, mostly through the work of anthropologists (E. Terray and M. Dobb). There is room, however, for a *theory of uneven development* under capitalism, encompassing international trade and imperialism. Some steps are being taken in this direction (see S. Amin and A. Emmanuel).

It is impossible to go into the details here. Instead, I sketch below some (limited) features of the general approach as applied to analysis of the problem which gave impetus to development of the modern theory of economic growth beginning with Harrod, that is, the problem of capitalist crises.

I. Surplus Value, Accumulation, and Capitalist Crises

The analysis utilizes a formal scheme the purpose of which is to constitute on an abstract and simple level the process of reproduction of the aggregate social capital in an expanding capitalist economy. Crises emerge as a result of the failure or inability of the system to satisfy completely the requirements of continued reproduction.[3] This failure is shown to derive from one or another of several *proximate* causes.

The starting point of the analysis is the process of production of value and surplus value under conditions of capitalism. Production consists of the production of commodities through employment of labor and means of production which are themselves the product of labor.[4]

From the value produced in a day's labor, a proportion, say ω, goes to sustain and reproduce the labor power of the worker. The rest constitutes surplus value or unpaid labor. ω is a socially determined magnitude. It is the labor value of the "necessaries" required to maintain the worker at a given standard of life as determined by historical and social conditions, including in those conditions the organized struggle of the workers vis-à-vis the capitalists, i.e., the conditions of the class struggle.[5] The rate of surplus value (or rate of exploitation), ϵ, is related to the value of labor power as

---

[3] This analysis does not deal with the mechanism of the business cycle, which represents the practical working out of the crisis, but with the broad conjunctures which may at one time or another underlie the existence of a crisis.

[4] The problem of fixed nonproduced resources (such as land) arises at a lower level of abstraction and, for purposes of a theory of accumulation, has to be dealt with in the context of an analysis of technical change which is being left out of account in this discussion.

$$\epsilon = (1 - \omega)/\omega \tag{1}$$

and it is the same in all sectors as long as ω is uniform.

Capitalists appropriate surplus value at the rate ϵ and competition dictates that the total amount of surplus value is redistributed among them at a uniform rate in proportion to the total price of their respective capitals. The rate of (net) profit, r, is thus the outcome of a market process of free exchange in which competition reigns because capital is freely mobile and prices of commodities are equated to their costs of production consisting of wages, plus profits, plus depreciation. In the equilibrium conditions of the price system, there exists a well-defined relationship between the rate of profit, the value of labor power, and the embodied-labor ratios corresponding to the given technique of production.[6] In the simple case where there are only two sectors of production (see Appendix), this relationship takes the form:

$$r = \rho(m_1, m_2, \omega). \tag{2}$$

Given the value of labor power ω and the embodied-labor ratios (m_1, m_2), the rate of profit consistent with the conditions of production and reproduction is uniquely determined from (2). The associated rate of exploitation is given by (1).

Consider now the requirement for overall balance of production and demand. In the aggregate, capitalists save a proportion s of their total profits.[7] Workers' saving is for consumption and their net saving as a class is zero. Investment is divided between increments to capital in each sector. For balance in the flow of output and expenditure it is required that savings equal investment. Thus,

$$sr = g = l_1 g_1 + l_2 g_2, \quad l_1 + l_2 = 1 \tag{3}$$

where g is the overall rate of accumulation, g_1, g_2 are the rates of accumulation in each sector and l_1, l_2 are appropriate weights. For the sake of simplicity we confine attention to the case of *balanced* growth where it is assumed that $g_1 = g_2 = g$.[8]

Next consider the investment plans of the capitalists. These may be represented as an increasing function of the expected rate of profit, r^e, and it is assumed for simplicity that the expected rate of profit is equal to the realized rate. Assume also that there is a minimum rate of profit below which capitalists do not invest. Thus,

---

[5] This is obviously not all there is to be said about the role of class struggle in the operation of the capitalist economy. The determination of the real consumption of workers is simply one point, and a significant one, at which it enters the system. Other relevant points in production are, for instance, the length of the working day and the technical conditions of the labor process. Further analysis would require introduction of these complications and recognition of the fact that, in practice, the wage bargain is conducted in terms of money wages while being tied to a cost of living index. When the workers' struggle is assumed to center around the share of wages in national income rather than around the real wage, the analysis would turn out rather differently (Harris 1972). It could be argued that the former is the case which is appropriate to the conditions of modern capitalism (Robinson 1973). What all of this points to is the need not only for abstract theoretical analysis but also for more concrete historical study of the labor-capital relation in modern capitalism.

[6] In the light of the work of P. Sraffa, this proposition can now be readily understood as one which holds in the *general* case of production of many commodities with or without joint production (in the form of fixed capital) and with one or many alternative techniques of production.

[7] This assumption is consistent with a more concrete formulation of the behavioral conditions in terms of corporations which retain a proportion α of total profits and distribute the rest to a rentier class who save at the rate s_r out of distributed profits. The saving rate of the whole capitalist class is then, $s = \alpha + s_r(1 - \alpha)$.

[8] There is, of course, no *empirical* basis for the assumption of balanced growth. Use of this assumption requires justification on other and quite different grounds. The rationale is strictly a *theoretical* one and is to be found in the specific focus of the analysis at this stage. In particular, the analysis is concerned with the reproduction and expansion of capital *as a whole*. For this purpose, the assumption of balanced growth is a convenient one. It leaves aside the problem of relative expansion of different sectors of capital, this being the purview of a theory of *uneven* development. For an equilibrium of balanced growth to exist, initial conditions must be appropriate. This requires a specific allocation of labor and means of production between the different sectors.

$$g = g(r^e), \qquad r^e = r. \tag{4}$$

This formulation is readily recognized to be a short-hand though fully adequate expression for a complex process.[9] It leaves aside the specific concrete conditions which may at one time or another govern the state of investment activity. For present purposes these may be regarded as causing shifts in the function $g(.)$.

The only remaining condition to be introduced is that concerning the availability of labor for production. Accumulation at any positive rate with the given technique of production sets up an increasing demand for labor. The requirement of an increasing labor force to match this demand is an independent condition of the problem.[10] Where the labor force grows at the rate ℓ this condition is

$$g = \ell. \tag{5}$$

These formal relationships constitute the basic structural conditions governing the expanded reproduction of capital. Consistency among them, taken together, is required for the process of reproduction and circulation to be smoothly carried out. Inconsistency among them accounts for a rupture or a crisis in the process. There are different possible types of crises and underlying causes, depending on the particular pattern of such inconsistency. These are illustrated with the aid of a diagram.

The left-hand quadrant of Figure 1 describes the relationship between the rate of profit and value of labor power for given technical conditions. At a given value of labor power, $\omega=\omega^*$, and associated rate of exploitation, the equilibrium rate of profit which allows competitive redistribution of the surplus among capitalists is $r=r^*$. In the right-hand quadrant the ray *OG* defines the relationship between the growth rate and profit rate which is consistent with overall balance of production and demand when capitalists save at the rate s. Call this the *profit realization curve*. It represents the profit rate that would be realized from sales of the product at a given rate of accumulation. It has a maximum at (r^*, g^*) corresponding to the existing rate of exploitation. The curves $g(r)$ represent the planned rate of accumulation as a function of the (expected) rate of profit. Three such

---

[9] On a social and historical plane, this process encompasses not only the inducement mechanisms of the market operating through obsolescence and technical change in methods of production, as well as the accelerator-multiplier effect of investment itself and the effect of wars, but also the social mechanisms determining entry and mobility within the capitalist class and various methods of "primitive" accumulation. While all of this is not capable of being captured in the simple formula of an investment function, it does require further articulation and analysis in theoretical terms. There is already a substantial basis for such an analysis in the works of Kalecki, Steindl, P. Sylos-Labini, and P. Baran and P. Sweezy, among others. Nevertheless, these more specific considerations are not strictly necessary at this stage of the analysis. This is again because of the particular level of abstraction at which the analysis is situated, that is, at the level of the reproduction and circulation of capital as a whole. At this level it is fully adequate to recognize that the drive for profits is the basic motive force underlying the expansion of capital. In this respect investment activity springs from the most fundamental characteristic of capitalist production and not from any arbitrary subjective or psychological motives (e.g., time preference) on the part of individuals.

[10] Expansion of the available labor force may take place in a variety of ways which are *internal* to the system of production as, for instance, through increase in the number of hours worked per worker or through absorption of the existing reserve army of labor viewed on a world scale (by means of immigration and capital export). It may take place also *on the margins* of the system of capitalist production as, for instance, through erosion of household work and other noncapitalist forms of production. In these and other ways accumulation creates the labor supply required for its own continuation. The connection is, however, neither fully automatic nor perfectly synchronized.

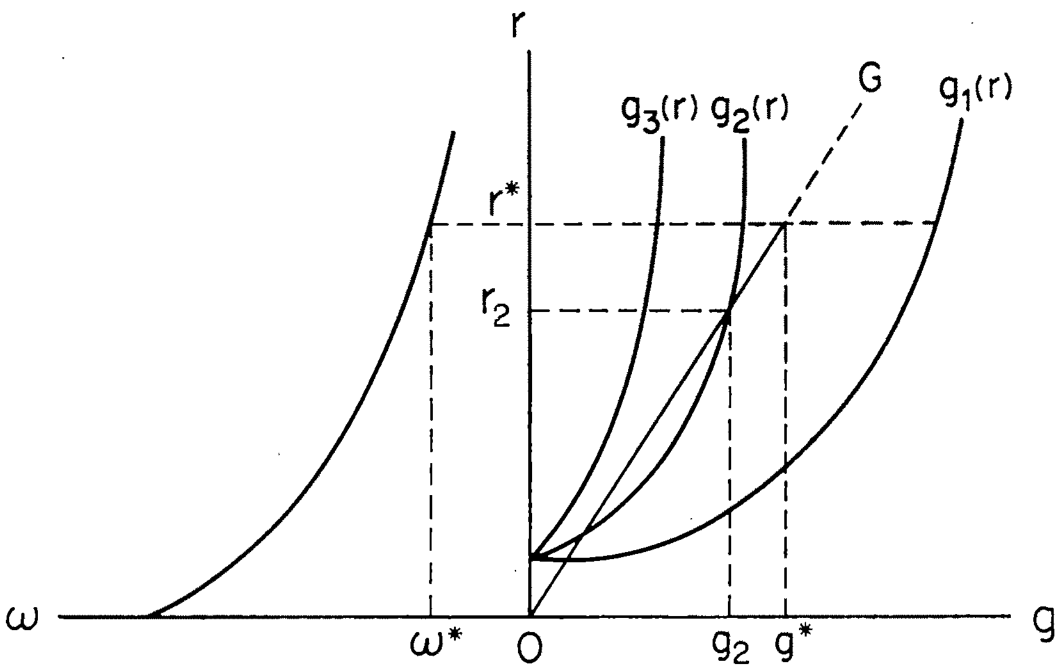


FIGURE 1

curves are drawn, each representing a different state of investment plans corresponding to a particular set of historical conditions. Consider the meaning of each case.

A. Case 1

When investment plans are at the level corresponding to $g_1(r)$, accumulation at any positive rate above some minimum enables the capitalists to realize a rate of profit which induces them to expand at a higher rate. The capitalists are seeking then to increase the rate of accumulation. With a higher rate of accumulation, a higher rate of profit would be realized and the capitalists would plan to grow at a still higher rate. There is, however, an upper limit to the rate of profit and rate of accumulation given by (r^*, g^*). At this limit the capitalists are, so to speak, straining at the bits to expand the size of their capital, $g_1(r^*) > g^*$. They are held in check by the rate of exploitation corresponding to the existing conditions. Under these conditions the planned rate of accumulation cannot be sustained. We may say that there exists a crisis. It is due in this case to excessive accumulation in relation to the pool of available surplus. It may be manifested, for instance, in the form of an inflationary spiral (compare Robinson's concept of an inflation barrier).

The actual rate of accumulation could be higher if the value of labor power were lower, thereby permitting a higher rate of profit. This would require a readjustment of existing class relations as related to the costs of reproduction of labor. Alteration of the technique of production, such as to lower the embodied-labor ratios and the labor embodied in wage goods, would have the same effect.[11] The actual rate of accumulation could also be higher if capitalists were to save at a higher rate. Otherwise, the only possible basis for resolving the crisis is through an adjustment in the rate of accumulation itself.

---

[11] It is the possibility of such an adjustment in production coefficients, viewed singularly as a matter of substitution along a given production function, which constitutes the central idea of the neoclassical parable.

B. Case 2

This is the case of investment plans corresponding to $g_2(r)$. Under such conditions, there exists a point (r_2, g_2) such that, when capitalists accumulate at the rate g_2, the realized rate of profit is just sufficient to induce them to continue to accumulate at the same rate. There exists an equilibrium, in this sense, with respect to the overall balance between production and demand. At the same time, however, the realized rate of profit is less than that which is allowed to them at the existing rate of exploitation, but the latter cannot be realized. Thus there exists a crisis. Call this a *realization crisis*. It may manifest itself, for instance, in a deflationary tendency (under competition) as prices fall below values and the rate of utilization of productive capacity declines below normal.

The realized rate of profit could be higher if the level of investment plans were higher than $g_2(r)$, or if the capitalists were to consume at a higher rate. In this sense, a realization crisis is due either to underinvestment or underconsumption on the part of capitalists, or to both. It is a matter of interpretation to which the analysis at this level is indifferent. Whatever the case, it is evident that there is no connection between the realized rate of profit and consumption on the part of workers. A higher level of workers' consumption leaves the realized rate of profit at the same level but *reduces* the rate at which capitalists as a whole can appropriate surplus value. In any event, this cannot come about without an adjustment in the prevailing class relations.

C. Case 3

With investment plans at the level corresponding to $g_3(r)$, the system is in a process of decelerating to a condition of zero growth with the realized rate of profit at its minimum level. There exists what we may call a *stagnation crisis*. It may be viewed as a degenerate case of a realization crisis.

D. Case 4

What if the state of investment plans were such as to correspond to a curve which passed through the point (r^*, g^*) from below? Then, at g^*, the realized rate of profit equals that which is appropriate to the existing rate of exploitation and is exactly what is required to maintain accumulation at the rate g^*. The situation need not, however, be fully consistent with equilibrium. This is because the requirement of a labor force to match the expanding demand for labor may not be fulfilled. The system may be in the course of exhausting the available reserve army of labor, being unable to replenish it sufficiently through internal or external means. Alternatively, the reserve army may be expanding because the rate of accumulation is too low to absorb the labor which is becoming available. Whatever the case, it is to be expected that each of these possibilities would, in turn, have effects on the rest of the system. These effects are likely to operate through variations in the bargaining power of workers and therefore on the rate of exploitation, with consequent feedback effects on the rate of accumulation. In this way the mechanism of operation of the reserve army of labor turns out

---

While it would be foolish to deny the technical feasibility of such adjustments, it is quite another matter to raise this to the level of a sufficient or even necessary condition for the resolution of a crisis. Because means of production are technically specific and particular patterns of organization of work and hierarchical relations of production are built around them in the factory, changes in methods of production are neither costless nor instantaneous. Neither are such changes a purely technical matter; they run up against the organized resistance of the workers.

to be crucial for the system as a whole, perhaps dominating every other form of crisis (Sweezy, p. 150).

Apart from the foregoing, there is more that can be said with this simple scheme. For instance, it is easy to recognize crises due to disproportionality between production of means of production and means of consumption. The role of monopoly, associated with different sectoral rates of profit, can be considered. To complicate the analysis, we have to introduce the conditions of technical change, the role of money, and so on. Further analysis would also require explicit recognition of the fact that the wage bargain is conducted in terms of *money wages*. The analytical difficulties are compounded when the problem of the state (an underdeveloped area of theory) is introduced. The objective of the analysis throughout is to provide historically based explanations as to why one or another set of crisis conditions comes to dominate the process of expansion of capital, to account for the mechanisms through which the crisis is resolved, and to explain the associated structural changes which occur in the system as it moves from one crisis to another.

Enough has been said here to indicate some broad features of the theoretical analysis. Meanwhile, yet another economic crisis is occurring in the capitalist world.

APPENDIX A

Let there be two produced commodities, each produced with one of them (number 1) and labor. The other (number 2) is consumed. There is a single, uniform period of production and stocks are used up in the period. For their maintenance and reproduction, workers consume a given quantity b per unit of labor. Labor values embodied in production of commodity i and labor power are

$$\lambda_i = a_{0i} + a_{1i}\lambda_1 = a_{0i} + A_{0i}, \qquad i = 1, 2$$

$$\omega = b\lambda_2$$

where a_{0i} are direct labor coefficients, a_{1i} are stocks and A_{0i} indirect labor. The rate of surplus value is

$$\epsilon = (\lambda_i - a_{0i}\omega - A_{0i})/a_{0i}\omega = (1 - \omega)/\omega.$$

In the price system:

$$p_i\lambda_i = (p_1 A_{0i} + p_2 a_{0i}\omega)(1 + r), \qquad i = 1, 2$$

where wages are paid out of capital and p_i are prices per unit of value. Setting $p_2 = 1$ and solving for r gives a quadratic from which, taking $r > 0$,

$$r = \rho(m_1, m_2, \omega), \qquad m_i = A_{0i}/a_{0i}.$$

REFERENCES

S. Amin, *Accumulation on a World Scale, 1, 2,* New York 1974.

P. Baran and P. Sweezy, *Monopoly Capital,* New York 1966.

E. Burmeister and A. R. Dobell, *Mathematical Theories of Economic Growth,* New York 1970.

M. Dobb et al., *The Transition from Feudalism to Capitalism,* New York 1954.

A. Emmanuel, *Unequal Exchange,* New York 1972.

R. A. Gordon and L. R. Klein, *Readings in Business Cycles,* Homewood 1965.

D. J. Harris, "On Marx's Scheme of Reproduction and Accumulation," *J. Pol. Econ.,* May/June 1972, *80,* 505–22.

———, *Theory of Growth and Distribution,* unpublished, 1973.

———, "A Post-Mortem on the Neoclassical 'Parable'," in E. J. Nell, ed., *Alternative Approaches to Economics,* Cambridge 1975.

J. Hirshleifer, *Investment, Interest and Capital,* Englewood Cliffs 1970.

E. K. Hunt and J. G. Schwartz, eds., *A Critique of Economic Theory,* Baltimore 1972.

N. Kaldor, "Capital Accumulation and Economic Growth," in F. A. Lutz and D. C. Hague, eds., *The Theory of Capital,* New York 1965, 177–222.

M. Kalecki, *Selected Essays on the Dynamics of the Capitalist Economy,* Cambridge 1971.

D. Levine, "Economic Studies: Contributions to the Critique of Economic Theory," unpublished, 1974.

E. Malinvaud, "The Analogy between Atemporal and Intertemporal Theories of Resource Allocation," *Rev. Econ. Stud.*, June 1961, *28*, 143–60.

K. Marx, *Capital, 1, 2, 3,* New York 1967.

J. Robinson, *The Accumulation of Capital,* Homewood 1956.

———, *Essays in the Theory of Economic Growth,* London 1962.

———, *Economic Heresies,* New York 1971.

———, "Ideology and Analysis," in *Sozialismus, Geschichte und Wirtschaft,* Festschrift für Eduard März, Vienna 1973.

R. Solow, *Growth Theory,* New York 1970.

P. Sraffa, *Production of Commodities by Means of Commodities,* Cambridge 1960.

J. Steindl, *Maturity and Stagnation in American Capitalism,* Oxford 1952.

P. Sweezy, *The Theory of Capitalist Development,* New York 1956.

P. Sylos-Labini, *Oligopoly and Technical Progress, Cambridge,* Mass., 1969.

E. Terray, *Marxism and "Primitive" Societies,* New York 1972.

K. Wicksell, *Lectures in Political Economy, 1,* New York 1938.

Growth Theory from an Evolutionary Perspective: The Differential Productivity Puzzle

By RICHARD R. NELSON and SIDNEY G. WINTER*

This is a further progress report on our efforts to develop a theoretical structure capable of integrating what is known about processes of technical change at the level of the firm or the individual invention and what is known about technological advance as reflected in industry or economy-wide time series data on output and inputs (see Nelson and Winter, 1973 and forthcoming). Our endeavor is motivated by a belief that a theoretical structure that integrates these two kinds of knowledge would be useful in enabling better understanding of certain important phenomena that have proved difficult to comprehend within traditional models. Prominent among these phenomena, and the focus of this paper, are the vast disparities in various measures of productivity growth and technical change across different sectors. To cite illustrative numbers (from J. Kendrick's 1973 study), the yearly change in output per worker (1948–1966) was 5.6 percent in farming, 4.6 percent in mining, 2.9 percent in manufacturing. Within manufacturing, yearly productivity growth ranged from 6.0 percent in chemicals to 1.7 percent in leather products.

The great intersectoral differences in growth experience clearly are a puzzle for growth theory. Nor is the puzzle strictly intellectual. While W. Baumol may have put the matter too emphatically, consideration of rapid productivity growth sectors and slow productivity growth sectors suggests strongly that many of the nation's problems relate to sectors where technological advance (using that term to serve as a proxy for productivity growth) has been slow.

There have been a number of attempts by economists to explain cross-industry differences within the framework provided by traditional neoclassical theory. Kendrick's 1961 study, particularly the work of N. Terleckyj contained in that study, set the style for much of this work. Roughly speaking, the technique has been to regress a measure of productivity growth against various industry variables, particularly research and development (*R&D*) spending of various sorts. Recent studies have broken *R&D* spending in an industry into self-financed and government-financed and have tried to account for the effect of *R&D* expenditures by supplying industries. The research has shown that industries differ greatly in the magnitudes of different kinds of research and development that feed into them and that these differences are related to differences in the growth of productivity. (For a good survey, see Terleckyj.)

The studies have been useful and provocative, but have not cut very deep. Clearly there are severe specification problems. The regression equations involve complex interactions in which factors influencing the demand side and the supply

---

* Yale University and the University of Michigan, respectively. Support for the project from the National Science Foundation is gratefully acknowledged, as is the helpful criticism on this paper by Susan Ackerman, Richard Levin, and Sharon Oster, all of Yale University.

side of the technical change process are intertwined and confounded with other forces. There is a tangle of causations, from *R&D* to productivity growth, from productivity growth and lowered prices to growth of output, from growth of output in the presence of scale economies to productivity growth, from expansion of the industry to greater incentives for research and development, and so on.

Even if it were granted that causation runs, at least in part, from research and development spending to enhanced productivity, what explains the great differences in (direct and indirect) *R&D* spending across industries? Among the range of possible explanations, two stand in stark contrast regarding their implications. One is that research and development activity is more powerful when directed toward the technologies of certain industries than toward the technologies of others; therefore, the disparities in rates of technical progress reflect some kind of innate differences in ability to advance efficiently the different kinds of technologies which *R&D* allocation decisions rationally reflect.

A second explanation (not mutually exclusive) focuses on differences in institutional structure that influence the impact of research and development on productivity and the extent to which *R&D* spending is optimal. The proposition is that the effectiveness with which *R&D* output is screened and employed varies among industries. Moreover, industries also differ significantly in the extent to which patent rights and other protections internalize the returns to inventions that flow out of firm *R&D* spending, and in the extent to which government subsidization permits the undertaking of high-valued research and development projects where externalities are important. Needless to say, the differences in these explanations matter profoundly in terms of their policy implications.

In order to probe the issues it would seem valuable to look beneath the industry time series data and analyze the more micro aspects of technical change in the different sectors: the sources of key inventions, their costs and who bore the costs, the motives of the organizations in the sector, the constraints under which they operate, how new technology is screened and spread, etc. Without arguing that it is impossible to enrich the basic neoclassical model so as to do this, for reasons discussed at length elsewhere (Nelson and Winter, forthcoming) we believe it more straightforward to work within a set of assumptions that in some ways narrow and in some ways broaden those contained within the more traditional model. We do not want to argue about whether our proposal really is a special case of traditional theory. Rather, the question we want considered is: Is this a fruitful way to theorize?

Our theoretical focus is on the following three questions: What are the sources of productivity growth in firms? What is the nature of the technological advance process? How can we characterize the way in which new technology is screened and spread throughout an economic sector? Put simply, we propose that the dominant source of productivity growth in a firm is new technology, that the processes that generate technological advance tend to operate stochastically and locally, and that selection and spreading forces in a sector work out rather slowly over time. In combination, these commitments represent the starting points of a "technological mutation" theory of changes in the characteristics of firms, with a "selection environment" operating on the new technological mutations—in short, an "evolutionary theory." But we have a lot more in mind than merely a metaphor.

Within an evolutionary theory, productivity growth is explained, in an accounting sense, in terms of, first, the generation of new technologies and, second, changes in the "weights" associated with the use of different existing technologies. We have been trying to model the stochastic processes that generate new technology in order to identify the factors that might explain why certain kinds of technical advances might be more likely than others. But the aspect of our work we want to sketch at more length here is the characterization of the selection environment for a new invention. Given a flow of new technologies, the selection environment determines how relative use of different technologies changes over time. The selection environment influences the productivity growth generated by any given invention, and it also influences strongly the kinds of *R&D* that firms in the industry will find profitable.

Consider an invention-innovation that has been introduced to the economy—for example, the first model 707 aircraft produced by the Boeing Aircraft Company, a new seed variety tried by a farmer, a pioneering doctor trying a new anticancer drug, the first use of "open classrooms" in a school system. A necessary condition for survival of an innovation is that it be perceived as worthwhile by the organizations that directly determine whether it is used or not. If the innovation is to persist and expand in use, the firm or farm must find a new product or process profitable to produce or employ, the doctor must view the treatment as efficacious, the school system must be persuaded that the new classroom technique is good educational practice and worth the cost. Sectors obviously differ greatly in terms of the nature of the organizations that ultimately must decide on worth and particularly the salient objectives that determine worth.

However, the question of whether or not the organizations (firms) find the innovations worthwhile depends not only on the objectives of the firms. In almost all economic sectors the firms—for-profit private organizations, public agencies, individual professionals—are subject to monitoring mechanisms that partly determine whether an innovation scores well or poorly according to the objectives of the firms and that may impose more direct constraints on firm behavior. A key part of this monitoring mechanism involves the individuals or organizations who are the demanders or beneficiaries of the goods or services produced by the firms in the sectors. Thus, the profitability of 707 aircraft to Boeing depends on how the airlines react to these planes. The seed must produce corn that is profitable at prevailing prices. Patients must agree to the new treatment, etc. In some sectors there are additional constraints imposed on the individual firms by agencies assigned a legal responsibility to monitor or regulate the activity of the industry. Thus the Boeing 707, before it could be put into commercial use, had to pass Federal Aviation Agency tests. New pharmaceuticals are regulated, etc. Selection environments differ greatly in the structure of demanders and monitors, and in ways their reactions feed back to influence the desirability of a particular innovation to the firms in question.

Industries differ not only in the objectives of the firms and demand and regulatory constraints facing them. They also differ in the relative importance of the mechanisms by which innovations, accepted by that environment, spread. There are, roughly speaking, two distinct kinds of mechanisms for the spread of an innovation. One of these is greater use of an innovation by the firm that first introduces it. If the firm produces a variety of products or undertakes a variety of activities, this may occur through substitution of the new activity for older ones. Or the firm may grow by investing new resources. In sectors which involve a number of administratively distinct organizational units on the supply side, there is a second innovation-spreading mechanism that needs to be considered—imitation. Imitation of certain innovations may be deliberately spurred by the institutional machinery: thus the agricultural extension service encourages widespread adoption by farmers of new seed varieties. Or the institutional machinery may deter or block imitation, as the patent system blocks the adoption by one firm of patented innovations created by a competitor. Industries clearly differ greatly in the importance of these two mechanisms and how they work.

We propose that a rigorous general model of the selection environment can be built from specification of these three elements: the definition of "worth" or profit that is operative for the firms in the sector, the manner in which consumer and regulatory preferences and rules influence what is profitable, and the investment and imitation processes that are involved. In the remainder of this paper we shall discuss some important qualitative differences in sectoral selection environments that become the focus of attention once one poses the theoretical problem in the way we have. Market sectors differ significantly among themselves. And many sectors involve important nonmarket components that have special characteristics. While we have not yet attempted modeling to capture the kind of differences in fine structure we shall be describing, this is high on our agenda of business.

The perception that market competition in a sector operates like a selection environment was explicit in the writings of many of the great nineteenth and early twentieth century economic theorists. J. A. Schumpeter was well within the classical tradition. In a stylized Schumpeterian evolutionary system, there is both a carrot and a stick to motivate firms to introduce "better" production methods or products. Better here has an unambiguous meaning: lower cost of production or a product that consumers are willing to buy at a price above cost. In either case the criterion boils down to higher monetary profit. Successful innovation leads to both higher profit for the innovator and to profitable investment opportunities. Thus profitable firms grow. In so doing they cut away the market for the noninnovators and reduce their profitability which, in turn, will force these firms to contract. Both the visible profits of the innovators and the losses experienced by the laggers stimulate the latter to try to imitate.

Both expansion of the innovator and imitation by competitors are essential aspects of the Schumpeterian process. In the standard descriptions of dynamic competition, expansion of the innovator is likely to be stressed. It is surprising, therefore, that the relationship between innovation and investment has not been subjected to more serious empirical investigation. While almost all theories of firm investment would lead one to expect a positive relationship between successful innovation and investment, this has not been tested in the major studies of investment behavior. The exceptions are studies where the author's basic hypothesis is oriented around the Schumpeterian interactions. D. Mueller does find that lagged *R&D* expenditure by a firm has a positive influence on its investment in new plant and equipment. E. Mansfield's studies (1968) also give support for a "Schumpeterian" view. Perhaps his most interesting results involve comparisons of firm growth rates, where he finds that innovating firms tend to grow more rapidly than the laggers. However, while the advantage of the innovators tends to persist for several periods, the advantage damps out with time, apparently because other firms have been able to come up with comparable or superior innovations.

In contrast to the sparseness of studies of the relationship of investment to successful innovation, there have been a large number of studies which have focused on the spread of innovation by imitation in profit-oriented sectors. It is not useful here to review the details of these studies, save to note that many of them have documented the role of profitability of an innovation in influencing the speed with which that innovation spreads. (For a good survey, see Mueller 1967.) In most of the cases studied, the innovations were inputs produced by a supplier and the early adopters were not in a position to block subsequent use by their competitors. In some instances this was not the case. For example, a glass producing company, Pilkington, holds the basic patents on the float glass process and presumably had an interest in limiting diffusion to other firms except where Pilkington was blocked from the market. It is interesting that the analysts of diffusion have not always been cognizant of these differences.

It also is quite surprising that in no study of which we are aware has there been an attempt to study the expansion of the innovator and imitation of the innovator together. It would seem apparent that, in order for a market selection environment to work effectively, a rather fine balance is required between the two mechanisms. Preservation of competitive structure in the face of innovation is delicate. A. Phillips' description of competition in the industry that produces aircraft for commercial airlines reveals one extreme. The "expansion of the successful innovator" mechanism works powerfully here. This is a sector in which firms are able to expand capacity rapidly and where it is costly and time consuming to imitate another company's successful product. Further, demanders are quite sensitive to product quality and cost. In this institutional regime, a company that comes up with a superior product has a great advantage over its competitors, and because of the lags and cost of imitation, other firms may be forced out of business. And indeed, it appears that successful innovation has come close to destroying competitive structure in the civilian air frame industry.

On the other hand, the efficacy of Schumpeterian competition requires that firms be able to achieve significant size and that imitation not be too easy. Consider the situation in agriculture where it is very difficult for a successful firm to expand significantly and rapidly because of the cost and complexity of purchasing particular pieces of adjacent land. This puts the full burden of the spread of profitable new technology on imitation mechanisms. But at the same time it limits the profitability to any individual farmer of innovating and thus reduces the incentives to search out new techniques and to take risks. Implicitly this has been recognized. An elaborate subsidized mechanism has developed to disseminate widely among farmers information regarding the best new techniques.

While economists have concentrated their attention on market sectors, research on the selection environments of nonmarket sectors has been undertaken principally by anthropologists, sociologists, and political scientists. This in itself would lead to some significant differences in focus and analysis. But to a considerable extent the differences in analysis reflect real differences in the selection environments.

A hallmark of nonmarket sectors is that the separation of interest between firms on the one hand and consumers and regulators on the other is not as sharply defined as in market sectors. The relationships between a public agency, like a school system and its clientele (students and parents), and sources of finance (mayor, council, and voters) simply do not have that arm-distance quality that marks the relationships between the seller and potential buyer of a new car. The public agency is expected to work in the public interest of its own volition and, indeed, to play a key role in the articulation of that interest. This is so in many nominally "private" sector activities, like the provision of medical services by a doctor. The doctor is not supposed to make his decisions regarding the use of a new drug on the basis of whether this will profit him, but rather on his expectation of how this will benefit his patients. This is not to say that, in fact, interests of firms and consumers always are consonant. In most nonmarket sectors (as in market sectors where competition is lax), the firm has a good deal of discretionary power regarding what it is to provide, and the customer may have little direct power to reward or punish performance.

In any case, the motivations of the firms in the sector are not assumed to be monetary profit. Analysis of the operative values of the firms, therefore, should be a centerpiece of studies of nonmarket selection. This makes analysis difficult. As in the theory of consumer behavior, as contrasted with the theory of the firm, tastes matter; these may be hard to analyze, and they may not be stable. Thus in the J. Coleman, E. Katz, and H. Menzel study of the diffusion among physicians of a new pharmaceutical, the authors did not even attempt to specify in what ways the new pharmaceutical was superior medically to preexisting alternatives. K. Warner's study of the diffusion of cancer chemotherapy reveals doctors trying to assess the patient's best interest, but often having few objective criteria to go on.

In sectors that involve significant regulatory or political controls or division of responsibility, there is an additional difficulty in analysis of the operational definition of worth. Many different parties may have to go along before an innovation can be operative. In L. Friedman's study of bail reform, the courts and the police both had to agree to the proposal, and legislative agreement was necessary where budget was involved. Thus collective decision processes are involved. The interaction of firms, and demanders and monitors, certainly is more subtle than in market sectors.

Nonseparation of suppliers and demanders leaves little room for firms to compete among each other for consumer dollars. Because of this innovations must spread largely through imitation across the spectrum of noncompetitive firms producing the goods or service in question. At the same time there is no incentive for the innovating firm to deter imitation. Organizations that cannot expand into the terrain of others and know that others cannot encroach on their territory have little to gain from preventing others from adopting their successful innovations. Indeed, in most of the sectors under consideration here, there are formal arrangements for cooperation and the flow of information among the firms.

Consider the quasi market for the provision of physician's services. Without strong constraints afforded by consumers or outside regulators, consumer welfare is guarded (perhaps not so securely) largely by professional standards of efficacy of treatment. To assess the efficacy of new treatments doctors consult with each other and apparently aim for professional consensus, guided by the judgments of certain key experts. L. Mohr's study of the spread of new practices and policies across local public health services reveals a similar "professionalism" at work. In such sectors, the screening and spreading of innovations depend in large part on these collective evaluation and information spreading mechanisms.

We hope we have persuaded you that there is considerable variation among sectors in their "selection environments" and that these differences can affect both the speed and the extent of spread of any innovation. It might be conjectured that these differences would influence the level of productivity at any time but not its rate of growth. We think this is wrong for two reasons. First, even if one assumes that the rate of advance of "best practice" is not influenced by the selection environment, it is not apparent that sectors need be characterized by a constant ratio of average to best practice. In some nonmarket sectors it is hard to identify strong forces that will prevent that ratio from falling. Second, and probably more important, the selection environment feeds back powerfully on the incentives for *R&D* by the firms in the sector. In industries like aviation there may be overstimulation. In agriculture, support of *R&D* is dependent upon suppliers or public sources of finance. Analysis of the incentives for *R&D* support by the firms in nonmarket sectors would appear to require a detailed study of operative values and sources of finance, but we would conjecture that in many cases the incentives for *R&D* support are quite weak. Thus one of the most important reasons for study of selection environments is that such analysis is necessary before one can evaluate the adequacy of prevailing *R&D* institutions.

Space limitations have forced us drastically to oversimplify and to curtail the discussion. We hope we have given at least a slight flavor of the kind of institutional richness that our proposed mode of theorizing aims to encompass. But the proof of the pudding will be in the eating.

REFERENCES

W. Baumol, "Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis," *Amer. Econ. Rev.*, June 1967, *57*, 415–26.

J. Coleman, E. Katz, and H. Menzel, "The Diffusion of an Innovation Among Physicians," *Sociometry*, Dec. 1957.

L. Friedman, *Innovation and Diffusion in Nonmarkets: Case Studies in Criminal Justice*, Yale Univ., doctoral dissertation, 1973.

H. Grabowski, and D. Mueller, "Managerial and Stockholder Welfare Models of Firm Expenditures," *Rev. of Econ. and Statist.*, Feb. 1972.

J. Kendrick, *Postwar Productivity Trends in the United States, 1948–1969*, Nat. Bur. of Econ. Res., Princeton 1973.

———, *Productivity Trends in the United States*, Nat. Bur. of Econ. Res., Princeton 1961.

E. Mansfield, *Industrial Research and Technological Innovation*, New York 1968.

———, "Determinants of the Speed of Application of New Technology," in B. R. Williams, ed., *Science and Technology in Economic Growth*, New York 1973.

L. Mohr, "Determinants of Innovation in Organizations," *Amer. Pol. Sci. Rev.*, 1969.

D. Mueller, "The Firm's Decision Process: An Econometric Investigation," *Quart. J. of Econ.*, Feb. 1967.

R. Nelson, and S. Winter, "Toward an Evolutionary Theory of Economic Capabilities," *Amer. Econ. Rev. Proc.*, May 1973, *63*, 440–49.

——— and ———, "Neoclassical versus Evolutionary Theories of Economic Growth," *Econ. J.*, forthcoming.

A. Phillips, *Technology and Market Structure*, Lexington 1973.

N. Terleckyj, "The Effect of R and D on the Productivity Growth of Industries," draft, Oct. 1973.

K. Warner, *Diffusion of Leukemia Chemotherapy*, Yale Univ., doctoral dissertation 1974.

CHINA AND INDIA: DEVELOPMENT DURING THE LAST 25 YEARS

Chinese and Indian Development: An Interdisciplinary Environmental Analysis

By BARRY RICHMAN*

This paper is based on my own firsthand research in the People's Republic of China and India during the second half of the 1960's, as well as information derived from numerous other sources. (See Richman and Richman and M. Copen, as well as the extensive documentation contained therein.) My interest here is in explaining major differences in the economic development of these two countries in terms of certain fundamental differences in their social development and underlying ideologies.

Economic development performance is reflected in such conventional indicators as gross national product and per capita income growth rates, industrial and agricultural output, etc. China has done considerably better than India in the last twenty-five years or so with regard to most indicators of economic development. (See, for example, the following works and the numerous sources cited therein: T. Weisskopf; K. I. Chen and J. S. Uppal; Richman; and Richman and Copen.)

I use the term social development here to include not only progress with regard to education and human resource development, health, welfare, social mobility, and raising real living standards, but also in connection with major changes in the attitudes, values, motivation, and behavior patterns of the population, as well as relationships between individuals, groups, and the broader society. Income distribution, which is also of major interest here, can be considered as part of economic and/or social development, although I treat it as being more a part of the latter since it tends to be more a sociological and ideological issue than an economic one.

I. Central Thesis

The central thesis of this paper is that the willingness and ability of people to work relatively hard and effectively on a national scale for goals that inspire them are perhaps the most important factors in explaining a country's economic development, and that such willingness and ability stem chiefly from a country's social development, as defined above. Moreover, ideology—or the lack of it—tends to play a fundamental underlying role in connection with the social development strategies and related performance pursued and realized by developing countries—in this particular case China and India.

My central thesis has evolved after developing and utilizing with some of my colleagues a conceptual framework of twenty-nine interrelated domestic or national environmental factors and fourteen international variables to explain significant differences in the economic development and management of different countries, including China and India. The environmental variables have been broken down into the following categories: educational; sociological-cultural (which includes psychological factors); political; legal; and economic. (See R. Farmer and Richman, Richman, and Richman and Copen.) Many of these factors are similar to the forty-one indicators of sociopolitical and economic organization and development that I. Adelman and C. T. Morris have focused on in their analysis of seventy-four countries (including India, but not China). Both their findings and ours have revealed an intimate association between the pace of economic progress and noneconomic forces relating directly to social development as defined here. This is also true of Gunnar Myrdal's massive, three-volume *Asian Drama* study.

---

* Professor of management and international business, University of California, Los Angeles.

I will focus on a number of aspects and indicators of social development which clearly seem to be associated with and highly relevant to explaining differences in the economic performance of these two countries. In fact, it appears that a good case can be made that there is a causative as well as a correlative relationship between various key aspects of social development and economic development performance. Moreover, it seems that social development strategies are often more important with regard to economic performance than economic strategies involving such things as type of economic system, resource allocation mechanisms, monetary and fiscal policies, organization of capital markets, the use of foreign assistance and resources, etc.

Before proceeding, a number of basic concepts and Western theories warrant brief comment, since they are relevant in explaining differences in Chinese and Indian development.

II. Some Basic Concepts and Relevant Theories

Harvey Liebenstein seems to have first coined the term X-efficiency and also presented evidence suggesting that other types of efficiency are often much more significant than allocative efficiency in terms of explaining economic growth and development. He has viewed motivation as being a major element in connection with X-efficiency, as well as other things, such as knowledge, information, and training. Motivation is critical in both bringing about and sustaining significant productivity or efficiency improvements and applying better methods and techniques on a broader front. Where motivation is relatively weak, there is likely to be considerably more slack in the system and hence a lesser degree of efficiency (see, for example, R. M. Cyert and J. G. March). Most human beings work at only a fraction of their overall potential in most societies, and gains from increasing X-efficiency are frequently much greater than those derived from increasing allocative efficiency.

The dominant attitude toward achievement—which relates directly to human motivation—and productive work tends to be a vital determinant of economic development (as well as social development). Studies pioneered by David McClelland and his colleagues clearly suggest that there is a rather close correlation between the achievement drive (or need for achievement) of different peoples and their countries' rates of economic growth. (See McClelland, 1961, 1963, and 1965;. McClelland and D. Winter; and Richman, pp. 281–91.) Achievement motivation is also related to social mobility, as well as attitudes toward education, training, and self-improvement.

Chester Barnard and Abraham Maslow have made pioneering contributions with regard to gaining a better understanding of human motivation by focusing on human needs and incentives. Barnard dealt with the "economy of incentives" and noted that material incentives seem to be rather ineffective positive motivators beyond the subsistence level for most people. Most human beings do not tend to work harder merely for material things, nor are they induced by material rewards to devote more than a fraction of their possible contribution at work. Barnard also considered "methods of persuasion" related to changing man's desires and attitudes with regard to human motivation, and this too is relevant to our central thesis.

Maslow presented five basic classes of human needs and placed them in a framework referred to as the hierarchy of needs. He hypothesized that, if all of a person's needs are unsatisfied at a particular time, satisfaction of the most dominant will be more pressing than the others. As fast as one set of needs is adequately satisfied, one tends to move up the hierarchy to the next set. This need-system or hierarchy is as follows, moving from the lowest to the highest level: (1) physiological; (2) safety; (3) belongingness or social needs; (4) esteem; (5) self-actualization. The last two are essentially psychological in nature.

More recent behavioral research by others has demonstrated more convincingly that material incentives tend to serve as necessary but frequently not sufficient conditions for productivity growth and efficiency over time. (See Richman and Farmer, especially chapters 14–16, and the sources cited in the bibliography.) Some job conditions may serve primarily to dissatisfy personnel, such as insufficient income, lack of job security, or poor working conditions. Yet improving these conditions often does not build strong motivation or contribute to significantly more effective and efficient performance. Social and psychological need-fulfillment tends to be more important in this regard.

Effective leadership and management over time usually involves having a good grasp of what results in individual and group satisfaction and dissatisfaction and how this relates to human motivation, behavior, and performance. Furthermore, truly effective and durable leaders and managers are usually perceived as being able to facilitate in significant ways the satisfaction of the personal needs and the attainment of the personal goals of their followers.

III. Ideology, Leadership, and Development

Ideologically and on the social front, China's basic aim has been to bring about a fundamental transformation of society, involving all of its social and economic institutions as well as the values, attitudes, motivation, and behavior patterns of its people. China's leadership has desired and striven for a collective egalitarian social system, a society without much of an individualistic orientation and with essentially public ownership of resources and capital assets—a nation of classless, cooperative, and altruistic citizens dedicated to the betterment of the broader society. And although Chinese society is still far from the ideal visualized by the more ardent Maoists, it is probably closer to this ideal than any other Communist country.

India's political leadership has sought to bring about major changes in traditional Indian society—and has been far less effective than China's leadership in this regard—but it has not rejected Western liberal values of individual freedom and private property rights. This involves a form of social democracy in which efforts are made to assure equal rights and social justice within the context of essentially bourgeois social, political, and economic institutions and to some extent within a mixed economy framework. In general, India is much further from its idealized society, and its ideology has been much more ambiguous and far less effective with regard to social change and overall development than has been the case in China.

The fundamental ideological differences of these two countries, and particularly how effectively their ideologies have been implemented, can explain much about the development strategies pursued by them and particularly about basic differences in social development which, in turn, relate to differences in economic development. For ideological reasons, China's leadership has pursued social development as an end in itself, rather than merely as a means to economic development, much more vigorously and effectively than India. As a consequence this has contributed much to China's superior economic performance.

Communist Chinese ideology consists of both the pure ideology of Marxism-Leninism and the practical ideology based on the "Thought" of Mao (see F. Schurmann for an excellent study of Chinese ideology). Pure ideology is viewed as theory based on universal truths dealing with values and ethics, and at its core is Marx's thesis of class struggle. It essentially involves a set of ideas designed to give the individual a unified and conscious view of the world. Practical ideology states norms and rules which prescribe behavior and have action consequences, but such ideology must be interpreted concretely to be operational. The Thought of Mao is a process of combining pure ideology or universal truths with concrete practice—practical ideology cannot be used until one has worked out correct thought, which does not simply mean a certain manner of thinking developed through intensive study, but also a manner of actually behaving. This approach has proven to be a very effective "method of persuasion" in China.

It is worth commenting briefly on a few of Mao's works that are most commonly studied and discussed in China. "On Contradictions" is studied in order to provide people with a "correct" and scientific view of reality and to help them in solving problems and overcoming obstacles. They are encouraged to persist in working out problems even where seemingly conflicting or incompatible factors are involved (e.g., raising product quality and reducing production costs simultaneously). "On Practice" is aimed at getting people to behave according to correct ideological standards through constant practice, experience, and evaluation. Another aim is to merge theory —including nonideological theories ranging from physics and engineering to bookkeeping and operating machines—with practice and application. Here, too, overcoming obstacles, problem solving, and persistence are stressed as well as self-improvement and better performance. "In Memory of Norman Bethune"—who was a Canadian surgeon who served with the Red Chinese Army in the field in the 1930's and died there—extols Bethune's virtues of selflessness, response to moral stimuli, altruistic dedication to his fellow man, classlessness, and humility. This work is studied with the aim of reducing the importance of self-interest and material gain as key motivating forces and of illustrating the value and correctness of a classless society mentality.

When the study of Mao's and other ideological works have not been carried to the extreme in China, they have undoubtedly had a major positive impact on China's social and economic development performance. China's party leadership has capitalized on, modified, created, and put into widespread application attitudes and values that are conducive to economic development—such as nationalism, the desire for achievement, work motivation, pride, dedication, loyalty, self-improvement, self-confidence, cooperation, resourcefulness, thriftiness, a more scientific view of the world among the masses, and so on. They have put these human forces into productive operation on a national scale through effective and persuasive—and at times outright coercive—political leadership, management, mobilization, and organization at both the macro and micro levels of society. The regime has relied on and effectively utilized the country's overall educational and training system, mass media, and entertainment and recreation of all types, as well as intensive study, indoctrination, discussion, self-criticism, and appraisal at all levels of society, typically under the direction of party or other ideological cadres.

In India, the shaping of an overall social environment conducive to more rapid and sustained economic development has been a far slower and much less effective process. This has been largely due to the lack of an effective underlying ideology, especially in terms of implementation, as well as the lack of effective political leadership, management, organization, motivation, and mobilization that are essential at both the macro and micro levels of society for major and basic social change on a national scale.

IV. Social Development Performance

The previous section focused on the importance of ideology with regard to development generally. At this point we shall examine a number of key aspects of Chinese and Indian social development performance which tend to support the central thesis of this paper.

A. Education and Human Resource Development

There tends to be a close correlation between a country's overall literacy rate and its level of economic development. The literacy rate is both a cause and result of development, and in a poor nation it seems to be one of the prime factors influencing development. In 1950 less than 10 percent of the Chinese population was literate compared to about 16 percent of India. In the 1970's about 50 percent of China's population is literate compared to less than 35 percent in India. Much of China's superior literacy performance is attributable to spare-time adult literacy programs, which have been relatively uncommon in India. (This section is based largely on data presented in Richman, ch. 3; Richman and Copen, ch. 3; and M. Blaugh, R. Layard, and M. Woodhall.)

In recent times roughly 40 percent more of China's population of primary school age has been enrolled in elementary schools than in India, and a much higher proportion has completed primary education in China. China has also provided a considerable amount of adult part-time primary educational programs, while India has had relatively few. The same has been true with regard to secondary education of all types. China also seems to have done a more effective job in imparting knowledge, skills, values, and attitudes in both primary and secondary education which are more conducive to economic development.

A significantly larger proportion of China's youths have been enrolled in and graduated from junior high school in recent times than in India. India has had a higher proportion in regular or general senior high schools, but the gap has been narrowing steadily. There has been much more stress in China's overall secondary educational system on technological, industrial, applied science, and agricultural training and on applied and practical education generally.

A major source of skilled and semiprofessional manpower for the Chinese economy is its system of specialized secondary and vocational training schools; and many of these also have part-time adult programs. India's overall system of this type lags far behind China in terms of enrollments, graduates, and probably also the quality and effectiveness of the instruction.

India has had larger enrollments and more graduates—in both total and per capita terms—in its overall system of higher education, but the overall growth of Chinese higher education has been significantly greater—with the notable exception of the Cultural Revolution period. China has also had much more part-time higher education and far fewer higher graduates unemployed than India. China has a substantial lead in technological, industrial, agricultural, and health fields and in the applied sciences—roughly twice as many students percentagewise have graduated in such fields in China compared to India in the past twenty-five years or so. India is significantly ahead quantitatively, and possibly also qualitatively, in the liberal arts, humanities, and the social sciences, though here too the gap has been narrowing steadily.

Hence, only in regular senior secondary and overall higher education have India's enrollment figures exceeded those of China. But this seems to clearly underline the relatively top-heavy structure of the Indian general educational system, which provides disproportionately great opportunities to those who have the resources to continue their education while at the same time limiting access considerably more than China to the bottom of the educational ladder.

Given the generally more favorable attitude toward practical and applied education in China, industrial, agricultural, and other personnel are more inclined to want to improve their skills, abilities, and performance through on-the-job training, self-education, and self-development, as well as through formal education, than in India. This has also been spurred on by the existence of many more opportunities for individual advancement in China.

B. Motivation

Two meaningful quantitative indicators of work motivation are labor productivity and absenteeism. The evidence available indicates that China has had significantly higher labor productivity in its economy generally and dramatically less personnel absenteeism as compared to India. (See Richman, pp. 260–61, 316, 514, 646–53, and 838–69; and Richman and Copen, pp. 532–35 and 590–91.) Some of the apparent important reasons for more effective human motivation overall in China as compared to India have already been noted. Here we shall elaborate and go into greater depth.

China clearly seems to have done a much more effective job of raising and capitalizing on the achievement drive of its population. China has succeeded in markedly raising the aspiration levels of its population in a variety of ways: formal education and training; effective ideological programs and campaigns, bringing about changes in child-rearing practices which are related to the breakdown and reorientation of the traditional family structure; and providing for greater educational and career opportunities generally. India has been considerably less effective overall.

India lacks a practical ideology of the Chinese type which transmits in standardized language concern with achievement on a continued basis to the studying and working population; and there has been no monolithic or effective political organization like the Chinese Communist Party which constantly stresses achievement and the virtues of hard work at the grass roots levels of society. Moreover, opportunities for upward social mobility and career advancement have generally been substantially more limited in India; this has been due largely to the class or social structure and slower rates of economic growth and expansion.

The fulfillment of social and psychological needs appears to have been emphasized to a much greater degree and extent and evidently with significantly greater effectiveness in China. Moreover, the physiological and safety needs of the common citizen have also probably been fulfilled to a significantly higher degree and extent. Apart from periods of ideological extremism, monetary incentives and personal gain have not been vigorously suppressed, but they have been restricted in their quantitative scope and supplemented by greater stress on moral stimuli, social and psychological need-fulfillment, and other incentives linked to a more egalitarian system of control over economic activity at the micro level. In other words, China seems to have been more effective in applying what Barnard has called "the economy of incentives," and this has resulted in more financial and other resources being available and used for reinvestment, economic expansion, and capital formation.

Three aspects of social development which are closely related to motivation warrant further attention: social mobility; income distribution; and participation.

C. *Social Mobility*

Educational and career opportunities are two of the most important indicators of social mobility in a country or society. The proportions of students from worker and peasant origins—and those who have actually been workers and peasants themselves—in Chinese secondary and higher education have increased very dramatically since 1949 and have comprised a significant majority of the overall student population at both levels for about fifteen years (Richman, pp. 297–98; and Richman and Copen, pp. 88–91). At the same time, except during periods of ideological extremism, students from less humble family backgrounds have also received higher educations in numbers equal to or greater than their proportionate representation in society. India's secondary and higher educational systems have been dramatically more elitist with regard to their student populations. As compared to China, vastly fewer students have been from worker and peasant families. This has no doubt contributed to a significantly greater willingness and ability for more people to work hard and effectively in China as compared to India.

My own study of the class origin of key personnel in a sample of thirty-eight Chinese industrial enterprises revealed that a substantial majority of such personnel were from worker and poor and middle peasant backgrounds (Richman, pp. 295–306). There have been many opportunities for individual advancement in Chinese industry and few barriers to promotion at most levels on the basis of ability—except during periods of ideological extremism when those from more traditionally elitist backgrounds have tended to be discriminated against, at least to some extent.

By contrast, studies of social structure in India underline the continuing significance of class and caste in limiting mobility and structuring the occupational hierarchy (N. Nijhawan; Richman, pp. 307–09; Richman and Copen, pp. 152–56; and the sources cited therein). Heredity, nepotism, family wealth and status, the "right social background" and "proper connections" have continued to play a large and pervasive role in determining opportunities for advancement in India. The vast differences in social mobility in connection with career advancement probably also contribute to the willingness and ability of a significantly greater proportion of China's population to work hard and effectively as compared to India.

D. *Income Distribution*

In India there has been little or no reduction in the highly unequal distribution that characterized that country at independence, and it is possible that the degree of income inequality has actually increased. (See Weisskopf, pp. 8–13, and the sources cited therein.) By contrast, China has been remarkably successful in narrowing income inequalities and thereby assuring a certain minimum standard to most if not all elements of the population.

My own survey of Chinese enterprises revealed that the ratio of top pay to average pay was typically no greater than 2.5 to 1. The highest ratio was 4 to 1. At most enterprises the differences between top and lowest pay was less than 4 to 1, and the highest paid employee in the entire sample received only about seven times more than the lowest paid. High-level company and plant managers in India often receive net incomes twenty or thirty times greater than the average worker (Richman, pp. 798–809). Moreover, the dramatic contrast between pay differentials in India and China actually understates the differences in the range of employee living standards. For in China many goods and services are provided more or less equally to all employees at highly subsidized rates. On the other hand, in India higher level executives often enjoy various job perquisites that are unavailable to lower level employees. Here the inequality in real income exceeds the inequality in money incomes.

The much greater income equality in China has served to satisfy human physiological and safety needs to a far greater extent than in India, and this has undoubtedly resulted in greater willingness and ability among many more people in China to work hard and effectively. It has also enabled greater and more effective use of moral stimuli and incentives linked to social and psychological need-satisfaction.

China's much more egalitarian strategy does call for greater diversion of resources from other uses to benefit the poor, and it may involve some dampening of individual material incentive gains. However, gains from such a strategy in terms of economic development in a poor country like China clearly seem to have outweighed the losses for reasons already noted. Greater equality in consumption can also contribute directly to improved worker health and productivity. Moreover, greater equality in income distribution can help limit wasteful consumption and the demand for scarce factors of production such as capital and foreign exchange (see Weisskopf, pp. 26–27).

E. *Participation*

China has emphasized employee participation in the management, planning, and decision making of organizations. A basic aim here, too, has been to bring about a more egalitarian society by reducing hierarchical and role distinctions. When not pushed too far too quickly, such participation has evidently contributed positively to the wllingness and ability of more people to work harder and more effectively. (See Richman, especially pp. 253–61.)

Employee participation can have positive effects with regard to the satisfaction of social and psychological needs. It can result in a greater sense of identification with and loyalty, belonging, and commitment to the organization. It can contribute to more effective education, training, communication, and information dissemination and feedback, since employees down the line analyze plans, problems, and results, exert initiative, and make suggestions for improving performance. Managers up the line can gain a better insight into the capabilities, attitudes, and opinions of lower level employees.

The degree and extent of employee participation in Indian productive organizations generally have been very minimal as compared to China. In general, Indian organizations typically utilize traditional hierarchical structures and authoritarian forms of management. There is also typically less decentralization or delegation of authority and much less voluntary cooperation, mutual assistance, and sharing of information in India as compared to Chinese organizations. (See Richman and Copen, chs. 4, 13, and 14.)

F. Health

Health is a very important factor in connection with the willingness and ability of people to work hard and effectively. Both India and China have made considerable strides in medicine and health care in the last twenty-five years or so. However, China's progress has been more rapid and effective, and China has been far more successful in spreading the benefits of improved health care throughout all levels of the population. (See R. Quinn; M. Wegman, T. Y. Lin, and E. Purcell; and Richman, pp. 538–40 and 551–55; and the sources cited therein.)

The much more equitable distribution and consumption of food in China also have a significant bearing on health. Even in absolute terms, China's per capita caloric food intake has been higher in recent times, and most people in China probably have had a more balanced and nutritious diet as compared to India.

V. Ideological Extremism and Chinese Development

References have been made at several points to ideological extremism in China, and some elaboration is in order here. Ideology has played a vital role with regard to China's impressive development; but when ideology has been pushed to extremes—in terms of economic, technical, and managerial feasibility or rationality—it has led to very serious economic (as well as political and social) crises (see Richman).

Thus far there have been two major periods which can be characterized as ones of ideological extremism in China. They were the Great Leap Forward of 1958–60, and the Great Proletarian Cultural Revolution of 1966–69. A basic ideological aim of both was to create rapidly a nation of dedicated and pure communist citizens as defined and perceived by Mao, or at least as defined in his name. This involves a nonbureaucratic society of selfless, altruistic, classless, and equal people.

During these two periods, the regime pursued vigorous campaigns, programs, policies, and practices with the aim of eliminating significant differences in the roles, status, privileges, material benefits, and living standards of leaders and followers, superiors and subordinates, managers and workers, generalists and specialists, party officials and the masses, the well-educated and the poorly educated, mental and physical labor, and so forth. A vigorous effort was also undertaken to replace material or monetary incentives and self-interest as key motivating forces with moral stimuli and appeals to altruism and self-sacrifice for the sake of the collective and the state. All of this had pervasive negative effects with regard to human motivation and need-fulfillment, as well as the effective utilization of knowledge and skills.

In an effort to implement the above programs the regime sought to have Reds (ideological purists) take over operating management and leadership functions normally performed by the experts (e.g., professional managers and administrators, engineers, scientists, technicians, staff specialists, etc.), because it was believed that the latter could not be counted on to implement such programs. When the Reds have taken over in this manner, ideology and politics have tended to take precedence over managerial, technical, and economic feasibility or rationality when a choice has had to be made—which typically has been frequently. Ideological indoctrination and political education sessions also became greatly intensified throughout the society—both during and after working hours. This too greatly constrained and inhibited the willingness and ability of people to work hard and effectively on a national scale.

It should be emphasized, however, that even during more normal periods of relative ideological moderation—and this has been most of this time since the Communists came to power in 1949—China has probably been the purest communist country in a relative—but by no means absolute—context. As already discussed, emphasis on ideology has had major and pervasive positive impacts on China's overall development.

VI. Conclusion

China has done considerably better with regard to both economic and social development than India has to date; and China has gone much further in creating an overall social environment conducive to sustained and substantial economic growth and development in the future. The apparent significantly greater willingness and ability of the Chinese people to work hard and effectively on a national scale for goals which inspire them may well be the most important factor in explaining China's superior economic development performance. Such willingness and ability stem chiefly from basic differences in the social development of the two countries, and ideology in conjunction with effective leadership clearly seems to have played the fundamental underlying role in this regard.

However, there appears to be an inherent conflict between the pure ideological goals and the ultimate economic, materialistic, and power objectives that Mao and his ardent supporters desire to implement and achieve: if China does evolve into a truly advanced economy, it would seem that some of the more important aspects of pure Maoist-Marxist-Leninist ideology would have to be compromised considerably in the process.

It is conceptually possible that the Chinese Maoists may have discovered a magic or spiritual tool which no other country has come upon. They may have a key which will allow ideology not only to create a pure philosophical state in reality but also develop a leading world productive system at the same time. To date, the available evidence indicates that the chances of this coming to pass are remote.

My own bet would be on some tradeoff between ideological aims and economic growth and development by China's current and future leadership. China may well be able to continue to be the purest communist state in existence—at least as conceived by the Maoists, though certainly not pure in the ultimate sense—and simultaneously achieve respectable and possibly even impressive economic growth and overall development over time. Unfortunately, the outlook for India's future development does not look nearly so bright at this time.

REFERENCES

I. Adelman and C. T. Morris, *Society, Politics, and Economic Development*, Baltimore 1967.

C. Barnard, *The Functions of the Executive*, Cambridge, Mass., 1938 and 1962.

M. Blaugh, R. Layard, and M. Woodhall, *The Causes of Graduate Unemployment in India*, London 1969.

K. I. Chen and J. S. Uppal, eds., *India and China: Studies in Comparative Development*, New York 1971.

R. M. Cyert and J. G. March, *A Behavioral Theory of the Firm*, Englewood Cliffs 1963.

R. Farmer and B. Richman, *Comparative Management and Economic Progress*, Homewood 1965 and Bloomington 1970.

H. Leibenstein, "Allocative Efficiency vs. X-Efficiency," *Amer. Econ. Rev.*, June 1966, *56*, 392–415.

A. Maslow, *Motivation and Personality*, New York 1954.

D. McClelland, *The Achieving Society*, Princeton 1961.

———, "Achievement Motivation Can Be Developed," *Harvard Bus. Rev.*, Nov.-Dec. 1965, *18*, 6–24.

———, "Motivational Patterns in Southeast Asia With Special Reference to the Chinese Case," *J. of Soc. Issues*, Jan. 1963, *19*, 6–19.

——— and D. Winter, *Motivating Economic Achievement*, New York 1969.

G. Myrdal, *Asian Drama*, New York 1968, *1, 2, 3*.

N. Nijhawan, "Inter-Generational Occupational Mobility," *Econ. and Polit. Weekly*, Bombay, Sept. 27, 1969, *4*.

J. Quinn, ed., *Medicine and Health in the People's Republic of China*, Washington 1972.

B. Richman, *Industrial Society in Communist China, A Firsthand Study of Chinese Economic Development and Management—With Significant Comparisons With Industry in India, the U.S.S.R., Japan, and the United States*, New York 1969 and 1972.

——— and M. Copen, *International Management and Economic Development: With Particular Reference to India*, New York 1972.

——— and R. Farmer, *Management and Organizations*, New York 1975.

F. Schurmann, *Ideology and Organization in Communist China*, Berkeley and Los Angeles 1966.

M. Wegman, T. Y. Lin, and E. Purcell, eds., *Public Health in the People's Republic of China*, New York 1973.

T. Weisskopf, *China and India: A Comparative Survey of Economic Development Performance*, dis. pap. 41, Cent. for Res. on Econ. Dev., Univ. of Michigan, Ann Arbor 1974.

China and India: Contrasting Experiences in Economic Development

By THOMAS E. WEISSKOPF*

Any effort to compare the development experiences of modern China and India must confront the fact that the development objectives of the political leadership of the two countries have differed in important ways. The Chinese Communist Party has clearly been aiming to bring about a fundamental transformation of Chinese society, involving all of its social and economic institutions as well as the attitudes and behavior patterns of the Chinese people. There have been internal differences over the extent and the pace of the change, but the overriding goal of the Chinese leadership has been to build a socialist—and ultimately a communist—society that rejects the individualistic orientation of most contemporary societies in favor of a collective and egalitarian social system. The Indian political leadership has sought to bring about major changes in traditional Indian society, but it has not rejected the Western liberal values of individual freedom and property rights. Instead, the Indian sociopolitical goals may fairly be summarized as the establishment of a form of social democracy, in which efforts are made to assure equal opportunity (but not equal outcomes) in the context of bourgeois political and economic institutions.

Within the framework of these overriding social objectives, there are three broad economic objectives that have been shared by both nations: rapid economic growth, greater economic equity, and greater economic self-reliance. I define economic equity to include not only an equitable distribution of income but also the provision of employment opportunities and basic social services for everyone. By economic self-reliance I mean the ability of nationals of a country—rather than foreigners—to make decisions that affect economic activity within the country.

I believe that on the whole the Chinese have been considerably more successful than the Indians in achieving the basic social goals of their development efforts. In many important ways the Chinese have succeeded in establishing more collective and egalitarian institutions and practices. India has maintained a parliamentary democracy and civil liberties to a degree that is exceptional among contemporary poor countries; but equal rights and social justice remain distant ideals. Unfortunately, such noneconomic aspects of development do not lend themselves readily to systematic comparison and evaluation. In the remainder of this paper I will therefore (reluctantly) confine my attention to the economic sphere of development.

My topic remains an extremely broad one, and I can only hope to provide a very cursory survey here. In Section I I will assess the progress of modern China and India with respect to the three broad economic objectives of development. In Section II I will contrast the Chinese and Indian approaches to development, and in Section III I will attempt to relate the differential progress of the two nations to differences in their basic approaches. I will conclude in Section IV with some observations about the underlying reasons for the differences in the development experiences of modern China and India.

---

* Associate professor of economics, University of Michigan. In writing this paper I have drawn both upon personal experience (four years of teaching and research in India in the 1960's and a brief visit to China in 1972) and upon the extensive professional literature on modern India and China. Rather than burden the text with continual footnotes and references, I have simply included the most important of my bibliographical sources in the list of references. The Twentieth Century Fund assisted in financing the research for this work; I am indebted also to the Center for Research on Economic Development at the University of Michigan for the use of its facilities.

I. Economic Development Progress

I have presented elsewhere (see Weisskopf 1974) a systematic and fully documented comparison of China's and India's economic progress in the modern postwar period. Here I will simply summarize the results of that study.

A. *Economic Growth*

The record of economic growth has unquestionably been better in China than in India. Most estimates of the average annual rate of growth of real output in China from 1952 (after an initial period of recovery) to the early 1970's range between 4 and 6 percent. The corresponding figure for India, based on official estimates, is 3½ percent. With population growing at about 2 percent per year in each country, the average annual rate of growth of *per capita* output has probably been twice as high in China (2–4 percent) as in India (1–1½ percent).

These long-term growth rates conceal considerable variation during the period under consideration. Of greatest current significance is the fact that since the mid-1960's the Chinese rate of growth has exceeded its two-decade average, while the Indian rate has been below the two-decade average. Moreover, in the past few years (beyond the range of the growth figures) there is evidence that the Chinese economy has fared much better than the Indian economy.

The more rapid growth of the Chinese economy from the early 1950's to the early 1970's was attributable primarily to the exceptionally rapid growth of Chinese industry. Chinese industrial growth averaged from 8–12 percent per year as opposed to 6 percent in India. On the other hand, Chinese agricultural growth may have been slightly less than that of India, averaging 2–3 percent per year as compared to roughly 3 percent in India. Once again, however, figures from the mid-1960's to the early 1970's and qualitative reports of more recent developments suggest that China's relative performance has improved in the past decade.

B. *Economic Equity*

It is in the area of equity that China's achievements have probably been the most impressive of all. Although there are no comprehensive data on the size distribution of income in China, all of the qualitative and fragmentary evidence suggests that there has been a vast redistribution of income from rich to poor since Liberation. My reading of the substantial (but often conflicting) quantitative information from India suggests that at best there has been no reduction in the highly unequal size distribution of income that characterized India at Independence, and most probably the degree of income inequality has increased. Limited evidence on urban-rural and interregional differentials in economic well-being suggest that both China and India may have narrowed some of the disparities, but that the improvement has most likely been greater in China than in India.

There can be no doubt that with respect to intraenterprise income differentials China has become vastly more egalitarian than India. Chinese differentials—in real terms—are probably the lowest in the world. In the provision of such basic social services as health care, literacy, and primary education, China's performance is also widely recognized as more successful than India's. India has made great strides in medicine and education, but China has made more progress in spreading the benefits of improved health care and education throughout all levels of the population.

Both China and India began their development efforts with a large burden of underemployment and unemployment. Unfortunately, there are formidable conceptual and empirical problems which make it very difficult to compare the changing employment situations in the two countries. The necessarily fragmentary information available points to the following tentative conclusions: (1) in urban areas overt unemployment is a much more serious problem in India than in China, for China has been more successful in absorbing labor into industrial and service activities (even if in relatively unproductive roles); (2) in rural areas underemployment remains a problem in both countries, but China's communal organization has been more effective in mobilizing rural labor and reducing the degree of underemployment than India's private farms and public agencies. Whatever the relative gravity of the problem of unemployment in China and India, there can be little doubt that China has managed much more successfully to provide the basic necessities of life to its economically marginal population.

C. *Economic Self-Reliance*

Although it is difficult to measure self-reliance directly, there are several indicators which are clearly relevant to its attainment. The most obvious of these is the extent of national control over productive assets. Since the early 1950's there has been no foreign ownership or control of economic activity in China. In India a substantial fraction of the modern corporate business sector (probably between 25 and 50 percent) is directly or indirectly under foreign control.

A second significant indicator of economic self-reliance is the extent to which a country is free of foreign debt. China received loans from the Soviet Union in the 1950's but repaid all of them by the early 1960's. Since then, China has maintained a generally favorable balance of payments and has remained free of any major foreign indebtedness. India has been plagued by a more or less continual foreign exchange crisis ever since the late 1950's and has had to make use of large amounts of foreign aid and loans to meet its balance of payments problems. As a result, India has found itself in a position of increasing indebtedness to the Western capitalist nations who have provided most of the capital inflow.

A final important indicator of economic self-reliance is the extent to which a country has developed an independent technological base for economic activity. This is inherently difficult to evaluate, but reports from qualified observers tend to agree that China is ahead of India in its overall technological capacity and particularly in its indigenous technological base. All in all, the evidence points clearly to the emergence of modern China as an economically more self-reliant nation than modern India.

II. Approaches to Economic Development

In both China and India there have been policy disputes and policy changes during the past quarter century, but the fundamental character of the approach to development in each country has remained fairly consistent over time. In this section I will briefly focus on some of the key differences in the Chinese and Indian approaches, as reflected in their basic economic institutions and overall development strategies.

A. *Economic Institutions*

The Chinese have systematically sought to minimize the role of private enterprise; they have relied almost entirely on public initiative and collective enterprise to spur economic development. Although various forms of private economic activity were tolerated during a period of transition in the 1950's, the pattern since the late 1950's has involved predominantly public control of both agricultural and nonagricultural activities. India has maintained a constitutional guarantee of private property rights and has accorded a predominant role to private enterprise in development. The Indian public sector has grown rapidly and includes many important modern industrial enterprises, but it still accounts for only a small portion of all nonagricultural activity and virtually no agricultural activity.

The Chinese economy is far more thoroughly planned than the Indian economy. Use of the market mechanism has varied over time and by sector, but resource mobilization and allocation have always been determined to a large extent by administrative means at various levels of the Chinese administrative hierarchy. In India, planning has in practice meant not replacement of the market mechanism but efforts to control it, either through expansion of government-controlled enterprise or through various forms of quantitative controls (e.g., investment licenses or foreign exchange allocations). Like private enterprise, the market has remained dominant in India, but it has had to contend with a great variety of interventions, restrictions, and distortions introduced by the government in its effort to gain some control over the economy.

In China serious efforts have been made to depart from the conventional mode of work organization, which involves a very hierarchical structure of authority within each enterprise and the use of individual gain incentives as a prime incentive to work. Although the Chinese experience shows periodic advances and retreats in this regard, individual gain incentives have been restricted in their scope and supplemented by collective incentives linked to a more egalitarian structure of control over economic activity at the micro level. In India no serious departure from the conventional mode of work organization has been contemplated in either the public or the private sector.

B. *Development Strategies*

Both the Chinese and the Indian political leaderships initially stressed the importance of institutional change for economic development, especially in the context of their predominantly traditional agrarian societies. In practice, however, the Chinese strategy has involved a much greater emphasis on institutional change. In progressing from early land redistribution to the ultimate collectivization of agriculture under a system of communes, the Chinese thoroughly transformed the character of their rural society. Indian efforts at land reform and "community development" have had much less impact on traditional agrarian social and economic relations. Disappointment with the pace of agricultural progress in both China and India led to a greater emphasis on technological change in agriculture beginning in the early 1960's. But in China the process of technological modernization has been linked to a new policy of rural industrialization, while in India it has been pursued without any major change in rural institutions.

In China the government—at various levels—has made a much more substantial effort to mobilize labor than in India. Apart from periodic mass campaigns engaging large numbers of workers in collective projects, the Chinese have systematically sought to organize communes, factories, neighborhoods, etc., so as to maximize employment opportunities. In India there has been some government subsidization of labor-intensive handicraft industries. But very few resources—financial or organizational—have been committed to large-scale public works projects or other employment-promoting activities.

The Chinese development strategy has placed a much greater emphasis on the welfare of the poor than the Indian strategy. Collective organizations developed in urban and rural areas in China have been designed to bring virtually everyone into a network of mutual social responsibility, whereas in India many people remain beyond the reach of public agencies and dependent on private charity.

Although in both countries the role of foreigners has varied over time, India's development strategy has on the whole accorded a much larger role to foreigners than China's strategy. Foreign private investment in China was liquidated soon after Liberation. The Indian Government has often spoken in economically nationalist terms, but it has in fact permitted a rapid growth of foreign private investment since Independence. China received economic and technical assistance from the Soviet Union in the 1950's, but since then has done without any net capital inflow and with only a very limited amount of (non-Soviet) foreign technical assistance. India received modest amounts of economic aid and technical assistance in the 1950's, but since the foreign exchange crisis of the late 1950's India has utilized a far greater amount of aid than China ever did. Throughout the 1960's foreign capital inflows—as well as foreign economic agencies—played a crucial role in India's development efforts.

III. Sources of Differential Progress

To what extent can China's greater progress in economic development be attributed to differences in the Chinese and Indian approaches to development, as opposed to differences in circumstances beyond the control of economic policymakers? With respect to the objectives of economic equity and self-reliance, differences in approach have clearly played a major role.

The greater success of the Chinese in overcoming inequalities and unemployment can be attributed in considerable part to the character of the basic economic institutions and development strategy espoused by the Chinese. First, the elimination of virtually all forms of individual private property in China removed the biggest source of income inequality. In India continued respect for private property rights has resulted in the perpetuation of enormous inequalities based on differential private holdings of land and capital assets. Second, the greater degree of centralized public control over resource allocation and lesser reliance on the market mechanism in China have permitted (but of course not necessitated) a more equitable allocation of resources among income classes and regions than has been possible in India, for the unimpeded operation of private enterprise in a market context tends to perpetuate—if not to exacerbate—initial inequalities. The deliberate Chinese emphasis on promoting the welfare of the poor and an institutional framework permitting major redistributive programs to be implemented have contrasted sharply with the somewhat sporadic concern for the poor shown by the Indian Government (depending on current political pressures) and an institutional framework rendering implementation of redistributive efforts very difficult. An additional aspect of the Chinese approach that has contributed to greater economic equity is the effort to limit the scope of individual material gain incentives; this has allowed for reductions in labor income differentials at the enterprise level to a degree inconceivable in India. Finally, the greater stress on the mobilization of labor in the Chinese approach—facilitated by an institutional framework involving strong public control over the labor market—has enabled China to combat unemployment and underemployment more successfully than India.

China's greater success in promoting economic self-reliance has resulted largely from deliberate strategic choices of the Chinese leadership. As a matter of policy, the Chinese decided to eliminate foreign investment while the Indians ultimately encouraged it in many areas. The Chinese deliberately pursued conservative international financial policies to avoid long-term indebtedness after the 1960's, while India welcomed large amounts of aid and loans which resulted in increasing debt. The Chinese allocated more resources and effort to the development of indigenous technological capacity, while the Indians relied to a much greater extent on the foreign technology introduced by foreign private enterprise. The termination of Soviet aid to China in the late 1950's and the lack of aid opportunities from other sources may have given an extra exogenous stimulus to Chinese efforts that India did not have. But in principle the Indian Government could have generated the same stimulus to self-reliance (at some possible cost in economic growth) by choosing as a matter of deliberate policy to curtail its utilization of foreign aid.

In the case of economic growth it is not so obvious that differences in the Chinese and Indian approaches to development are responsible for the differential performance. It is conceivable that uncontrollable differences—e.g., in cultural heritage, in ethnological structure, in initial economic conditions, in natural or external events—may have favored more rapid growth in China than in India. However, I believe that there are several important respects in which the Chinese approach to development has been more conducive to economic growth than the Indian approach.

First of all, the Chinese approach has involved a much more radical transformation of social and economic institutions than the Indian approach. In virtually all spheres of society the Chinese have brought about fundamental changes that must have profoundly affected the values, attitudes, and behavior patterns of the Chinese people. In many cases the institutional changes may have been so rapid or so unsettling as to be economically counterproductive, at least in the short run. Periodic changes and even reversals of policy suggest that this has often been true. But the radical institutional changes brought about in China seem most likely to have had the long-run effect of creating a social and psychological environment far more conducive to economic and technological progress than existed in prerevolutionary China or than exists in contemporary India. With the help of a strong governmental and party apparatus and a revolutionary ideology, the Chinese leadership has managed to reach virtually all of the people and to organize and motivate them to economic achievement. In India the scope of institutional change has been rather limited and—as a consequence—the plans and the policies of the Indian Government have had relatively little impact on the lives and the attitudes of the vast majority of the Indian people.

A second major aspect of the Chinese approach to development contributing to the relatively impressive growth performance of modern China has been the strong control over the economy exercised by central authorities. The dominant position of state planners in China contrasts with a complex diffusion of power in India among government planners, private businesses, and various other interest groups. Although the Chinese Government has moved in the direction of increasing decentralization of economic decision making within the overall administrative framework, the central planners have always been able to ensure a high rate of surplus mobilization and utilization for developmental purposes. Luxury consumption has been virtually prohibited, and most consumption has been kept within essential and productive forms—i.e., consumption that improves the health, education, and general productivity of the labor force. Indian economic policy has been marked by conflict and compromise among competing interests, with no single group exercising decisive authority. On the one hand, powerful private groups have prevented Indian planners from limiting nonessential and unproductive consumption and from raising public resources and allocating them for development purposes in the amount and direction desired. On the other hand, political pressures from middle-income classes have periodically limited the responsiveness of Indian Government policy to big business interests and brought it into conflict with the desire of domestic and foreign capitalists for relatively unfettered profit-motivated expansion. The end result has been an economic strategy involving a mixture of public and private authority that has interfered with both socialist and capitalist accumulation.

The third characteristic of the Chinese approach to development that would appear to have favored faster economic growth than India's is its greater egalitarianism. No doubt an egalitarian strategy does call for the diversion of resources and energies from other areas to benefit the poor, and it may involve some dampening of individual economic gain incentives. But there are also many respects in which more equality can be functional for growth, and these respects appear especially likely to dominate in very poor and initially very unequal societies such as those of China and India. First, greater equality in consumption can contribute directly to improved worker health and productivity. Second, a more equal economic environment is likely to increase the work motivation of the labor force—particularly in the case of public works projects employing surplus labor. Finally, greater equality in income distribution may help to limit wasteful consumption and the demand for scarce factors of production such as capital and foreign exchange.

There are some respects in which the Indian approach may have been more favorable to growth than the Chinese approach. As noted earlier, India has made much greater use of foreign aid, investment, and technology than China. This has probably contributed to faster economic growth in the short run, although it is arguable that in the long run extensive utilization of foreign resources may be counterproductive because—among other things—it can inhibit the development of indigenous capabilities. The greater role accorded to the market mechanism in the Indian approach might also be considered a source of greater efficiency and hence more rapid growth in India than in China. However, the Indians as well as the Chinese have interfered in many ways with the operation of their internal and external markets. And—neoclassical orthodoxy notwithstanding—it is doubtful that pure allocational efficiency really has a significant impact on economic growth in poor countries such as China and India.

In sum, neither the Chinese nor the Indian approach to development can be considered optimal with respect to the achievement of rapid economic growth. Yet the Chinese approach appears to have been significantly more conducive to growth as well as to equity and to self-reliance than the Indian approach.

IV. Conclusion

Underlying the differing Chinese and Indian approaches to development are significant differences in the political-economic conditions that have characterized China since Liberation and India since Independence. The most crucial of these differences seem to me to involve the distribution of political power and the authority of the political leadership.

In India the attainment of Independence brought about no major redistribution of power between previously dominant and dominated classes. There was some shift of power among elites: indigenous businessmen and professionals gained at the expense of their foreign counterparts, and some of the more enterprising large landholders gained at the expense of traditional rural potentates and absentee landlords most closely linked to the British. India's parliamentary democratic system has enabled certain intermediate groups—such as moderately well-to-do peasants and organized urban workers—to have some voice in national affairs, but the vast majority of poor peasants, landless laborers, and unorganized urban dwellers remain as powerless as ever before. In China, by contrast, a truly revolutionary redistribution of power did occur as a result of the Communist Liberation. All of the old privileged classes—foreign and domestic, urban and rural—saw their power sharply curtailed (if not totally eliminated) as the Chinese Communist Party took full control of the nation. The extent to which the previously oppressed classes are now able to exercise real power is a matter of much controversy. But it is significant that the political leadership is fully committed ideologically to the supremacy of workers and peasants, and that failure to act in their interest would entail a heavy political—and probably economic—cost.

In India the capacity of the national political leadership for decisive government action is limited by several important political and institutional factors. First of all, the system of parliamentary democracy—for all its inability to equalize the distribution of political power—does permit certain interests to mount opposition to various aspects of government policy and thus to constrain the role of the central government. Second, the federal structure of government in India (in part the result of India's ethnological heterogeneity) allocates certain critical responsibilities to state governments and thereby weakens the power of the central government and the national political leadership. Third, the extensive protection of private property rights guaranteed by the Indian constitution (reflecting the interests of powerful propertied classes) sets strict limits on the authority of the state to bring about institutional changes affecting private property. Again, the contrast with China is striking. The Chinese political leadership has the authority and the capacity to exercise decisive control over Chinese society and economy. Through well-organized structures of party, army, and civilian administration, the Chinese leadership is able to implement its chosen policies with far greater effectiveness and thoroughness than the Indian leadership.

It is easy to see how the differing distributions of political power and the differing degrees of central authority have led to most of the differences in the Chinese and Indian approaches to development. In India the strength of private propertied classes has limited the role of the public sector; the weakness of central government authority has impeded the capacity for national planning; the very unequal distribution of power has inhibited the implementation of radical social change and egalitarian policies; and the limited capacity of a weak central government to mobilize the economic surplus, together with the continuing strength of private capital, has led to a considerable role for foreigners in India's development. In China, the elimination of the formerly privileged classes and the authority of the new leadership have made possible radical changes in social and economic institutions, involving a dominant role for public and collective enterprise and the effective implementation of egalitarian and nationalist policies.

A review of the political economy of the development experiences of China and India suggests that India has been paying not just the price of democracy (in a weakened central authority) but also the price of capitalism (in an inegalitarian property-dominated distribution of power). It is arguable that with a strong central authority devoted to unfettered capitalist expansion, India could have achieved a higher rate of economic growth over the past two decades. By removing restrictions on domestic and foreign capitalists, suppressing the consumption demands of the middle and organized working classes, and abandoning even the limited redistributional efforts of the past, a strong Indian government might have been able to create conditions much more favorable to private capitalist accumulation. But at best such a framework would have promoted a very uneven and politically tenuous form of economic development. The Chinese experience suggests that a pattern of development combining economic growth with an increasingly equitable and self-reliant society requires not only a stronger central authority but also a dramatic shift of political power away from those classes who benefit from inequality and foreign involvement. Without such a fundamental transformation of political-economic conditions in India, it is difficult to be optimistic about India's prospects for attaining the several economic objectives of development.

REFERENCES

C. Bettelheim, *India Independent,* New York 1968.

J. Bhagwati and P. Desai, *India: Planning for Industrialization,* London 1970.

K. I. Chen and J. S. Uppal, eds., *India and China: Studies in Comparative Development,* New York 1971.

A. Eckstein, W. Galenson, and T. C. Liu, eds., *Economic Trends in Communist China,* Chicago 1968.

A. H. Hanson, *The Process of Planning: A Study of India's Five-Year Plans, 1950–1964,* London 1966.

A. Maddison, *Class Structure and Economic Growth: India and Pakistan Since the Moghuls,* New York 1971.

G. Myrdal, *Asian Drama,* New York 1968.

J. S. Prybyla, *The Political Economy of Communist China,* Scranton 1970.

B. Richman, *Industrial Society in Communist China,* New York 1969.

T. E. Weisskopf, "Dependence and Imperialism in India," *Rev. Rad. Polit. Econ.,* Spring 1973, *5,* 53–96.

———, "China and India: A Comparative Survey of Economic Development Performance," dis. pap. no. 41, Cent. for Res. on Econ. Development, Univ. of Michigan, Oct. 1974.

E. L. Wheelwright and B. McFarlane, *The Chinese Road to Socialism,* New York 1970.

Joint Economic Committee, U. S. Congress, *An Economic Profile of Mainland China,* Washington 1967.

———, *People's Republic of China: An Economic Assessment,* Washington 1972.

DISCUSSION

Padma Desai, Harvard University: I have undertaken the task of commenting on the two papers by Barry Richman and Thomas Weisskopf with considerable hesitation. Coming from an Indian, my assessment of the contents and especially the conclusions of these papers is likely to be suspect. I shall try my best to avoid the temptation of being less than impartial; and I hope that any prejudice on my part will not be prejudged.

I also happen to have some reservations about comparing and contrasting the problems and performance of these two countries. There are, of course, similarities between the two in terms of their enormous land and population masses and the historical burden of long foreign exploitation. However, each has been operating under a different political system; besides, their record of economic growth is so brief as to be more an experiment defying attempts at prediction rather than a sustained experience. Moreover, the problems of each are staggering. Perhaps my familiarity with the mind-boggling Indian scenario breeds some caution. The Chinese scene is equally complex with the additional handicap of lack of reliable information on China.

With these introductory remarks, I proceed to comment on how successfully these courageous men accomplish the task of contrasting Chinese and Indian economic performance and explaining the sources of their differential achievements.

I begin with Richman's paper. He provides a conceptual framework for explaining the better economic performance of China in terms of its superior social development. The sequence of argument proceeds as follows: the higher the level of social development, the greater the willingness and ability of the people to work "hard and effectively on a national scale for goals that inspire them"; as a consequence, economic growth is more rapid.

Social development, in turn, can be identified in terms of progress not only in education, health, social mobility, and so on but also in the attitudes, values, motivation, and behavior patterns of the population. Thus Richman states: "China's party leadership has capitalized on, modified, created, and put into widespread application attitudes and values that are conducive to economic development—such as nationalism, the desire for achievement, work motivation, pride, dedication, loyalty, self-improvement, self-confidence, altruism, cooperation, resourcefulness, thriftiness, frugality, a more scientific view of the world among the masses, and so on."

Furthermore, China's leaders have sought to create such a perfect state of social development by a massive and sustained employment of communist ideology and organization.

Obviously it is impossible to test this qualitative framework rigorously. Instead, Richman picks out certain criteria of social development, such as level of education, work motivation, social mobility, income equality, employee participation, standard of health, etc., and then identifies them quantitatively for both China and India. For example, high labor productivity and low absenteeism indicate superior motivation; then again, educational and career opportunities are suitable indexes of social mobility and so on.

On the basis of tenuous observations, he concludes that on almost all counts of social development China has done better than India; therefore, by inference from his model, Chinese economic performance is superior to Indian; and he proceeds to assert that there are quite a few economic indicators which support this claim.

I have reservations both about the analytical framework and its actual application to China.

To begin with the conceptual shortcomings of the framework: a heightened level of social awareness with its attributes can lead to faster economic growth if its total contribution to the growth process is positive. Richman does not, however, provide us with the dynamics of the interaction from social development to economic growth and so I am left worrying about the negative contribution to economic growth of some of the elements associated with a state of social awareness. For example, there is a critical point at which ideological exhortation and enforced discipline in a society such as China begin to contribute negatively to growth. Also, in traditionally hierarchical societies such as China and India, a clear demarcation of authority rather than an egalitarian participation at each stage of the work process may be more conducive to productivity.

Richman's own awareness of the tradeoffs between enforced egalitarianism and productivity and between ideological extremism and economic growth comes through almost as a letdown on the concluding page of the paper when he declares: "If China does evolve into a truly advanced economy, it would seem that some of the more important aspects of pure Maoist-Marxist-Leninist ideology would have to be compromised considerably in the process."

As for the actual application of the framework to China, the Chinese record of the last twenty-five years gives evidence of the contradictions between a presumably superior societal state and a productive economy. In other words, the Chinese leaders seem to be less certain of the positive link between social development and economic growth. As a result, we have, as Richman points out in a brief section, the phenomenon of the excessive exuberance of the Great Leap Forward and the ideological excesses of the Cultural Revolution followed by some economic rationality.

I have also some reservations about the tenuous evidence in terms of which the author argues that China is a superior societal state and hence a more productive economy than India. I shall comment on that aspect in my concluding remarks.

I now go on to the second paper by Weisskopf. He argues that China has performed better than India on all the three counts of economic growth, economic equity and self-reliance in the recent period.

He attributes the differential performance in self-reliance in large measure to the differences in their basic strategies of development. It is argued that the Chinese system is more self-reliant because it has not relied to an equal degree as India on foreign private investment, foreign aid, imported technology, and so on. It is further argued that the Indian reliance on foreign private investment has led to economic dependence on foreigners. I see here a ready willingness to equate India with a South American banana republic. In this context, it is possible to argue that given India's size, resource diversity, and political sophistication, India has been quite aware that she could absorb more private foreign investment without compromising economic or political independence. After all, as we Indians like to say, one cannot walk with an elephant in one's pocket.

It is also arguable that the Indian strategy of utilizing aid and limited private investment to finance imports of technology in order to narrow the technology gap has produced far more substantial and variegated technological absorption in India than has been the case with China: thus witness the recent report of Norman Borlaug on the state of agricultural research in the two countries. Finally, Weisskopf must not forget that the Chinese tried to use Soviet aid and were *forced* to do without it only because the Soviet Union withdrew abruptly from the scene: the wisdom attributed by Weisskopf to the Chinese was really Hobson's choice.

As for the superior Chinese performance in economic equity, Weisskopf traces it to the Chinese strategy of the abolition of private property, creation of mass employment, and extensive provision of social services. It seems to me that some societies which have abolished private property have ended up by creating other types of inegalitarian privileges, such as those arising from bureaucratic abuses, whereas other societies which have retained private property also are characterized by varying degrees of economic inequality. Thus it would seem that the abolition of private property in itself does not necessarily result in greater equality.

Nor is it accurate to exaggerate the prevalence of equality in the Chinese countryside as measured by the provision of employment and other services. What the Chinese seem to have accomplished is widespread distribution of food and the most rudimentary types of shelter, education, and medicine to its population in a commune-type organizational system in which free mobility is prohibited. However, *within* this overall arrangement, there must be differences in the levels of remuneration, in the quality of social services provided from commune to commune, from village to village, from one factory to another, and from some favored regions on the Eastern seaboard to the outlying wilderness of Tibet. I refuse to believe that if one has seen one commune or factory, one has, like Spiro Agnew, seen them all.

In the concluding section, the Weisskopf paper attempts an explanation of the better economic growth performance of China. It is argued that China has grown faster as a result of its social and ideological transformation, its authoritarian political system, and its egalitarian strategies. This section of the paper resembles Richman's framework which seeks to relate economic growth to superior social and ideological development. I have already expressed my reservations about this framework earlier.

My overriding hesitation about both the papers is essentially methodological. Some of us who have been working on India for quite a few years find it baffling to provide decisive answers to issues such as how self-reliant or egalitarian or technically mature India happens to be in spite of the mass of information and extensive access to it by scholars for scientific scrutiny and evaluation—unlike in China. Similarly, the bits and pieces of information which are allowed to trickle in from China are certainly inadequate to provide answers to similar issues in the Chinese context, notwithstanding the assertions of China-watchers from Mainland America.

Jerome Cohen has recently characterized the blinkered activity of several Sinologists as "Marcopolities." I am reminded of a funny story by Stanley Karnow who stayed on to visit Peking University after the Nixon entourage had left. Out of nowhere, an apparently random sample of Chinese students appeared to entertain and inform him. Karnow was congratulating himself on his success in meeting this unbiased group until one of these students said: "As I was telling James Reston last week . . . ".

When facts are scarce and scholarly analysis and scrutiny are rare because infeasible, how does one arrive at the critical judgments that are necessary to evaluate and compare the Chinese society and economy with the Indian? And so I ask myself: How can one be so sure, as Richman, that Chinese managers are good and Indian managers are incompetent? Perhaps there are good managers and bad managers in both systems. How can one assert, as both papers do, that China is technologically more advanced than India? Where is the systematic and detailed documentation of the research activities in the industrial, agricultural, and educational institutes and farms and factories in the two countries? Is it really true that China has solved its unemployment problem? What does such a statement mean except to imply that labor in China is not free to move where it chooses to? I am still at a loss to decide if the Chinese promote productivity by rewarding merit or via ideological devotion. How does the whole system of rewards and penalties, which is the essence of productive activity in any society, really function in China?

I trust, therefore, that I will be excused if I remain on the fence, waiting for the all-too-missing evidence and careful analysis of the Chinese experiment before I accept the conclusion that it Pareto-dominates the Indian economy on all dimensions. I cannot help recalling listening in early 1967 at Columbia University to a noted Sinologist of an ideologically quite different persuasion from the authors of today's session. He was favorably comparing the Indian economy with the Chinese, using the Indian Plan-document statistics, representing mainly *ex ante* aspirations, and Chinese *ex post* performance. I have a distinct impression that we have the reverse phenomenon in the two papers today: *ex ante* Chinese (and Sinological) assertions being compared with *ex post* Indian realities.

China will have to be opened up by the Chinese to far more objective scrutiny and the suspension of disbelief will have to be renounced by the China-watchers before a realistic assessment can be expected on the issues that both the papers today claim to have settled but have only barely raised. This may well take more than a generation: a time scale which is atomistic in India's and China's civilizations but which is unhappily too long for the American. One must not also forget that much can, and historically often does, happen in two countries with such diverse and difficult problems of domestic management and international relations when one shifts to a time scale rather longer than that implicit in the papers today; and, indeed, little of such prospective performance can be captured by our analysis even if we supplemented our economic apparatus with historical and sociological analysis. I like to recall Selig Harrison's disturbing forecast of India's political balkanization and disintegration in his well-known book, *India: The Most Dangerous Decades*, and how, when the decade and more had passed, a witty friend of mine from Tamil Nadu in India approached him at a Washington party and said: "We tried hard to implement the program you had laid down for us, but you know how inefficient we are." And I have a matching story for China too. As Weisskopf certainly knows, unlike the subjects of their study, the Sinologists and the Sovietologists at Harvard sup at the same subsidized table in peaceful coexistence. As I stood one day in the line behind two Sinologist friends in animated discussion of Chinese politics, the possible fragility of our predictions regarding China's political and economic future dawned on me as I realized that one was a reactionary Sinologist if one wished Mao Tse-tung to die before Chou En-lai and a revolutionary Sinologist if one wished the opposite.

Rather than conclude on these skeptical thoughts, let me raise two questions to leave with the authors for their further reflection. First, given the totally different political systems which India and China inherited and possess, for historical reasons which neither can be really given credit or discredit for, would it not be a more meaningful question to ask whether each is efficient or equitable *within the constraints* specified by the political system? This is surely a necessary complement to the question normally posed, as in today's papers: i.e., whether one *overall system* is superior to the others on efficiency or equity grounds. Next, even if the overall Chinese system were Pareto-superior on desired dimensions, what does it really imply about the critical choices before other developing countries? And I mean here not merely the predicament of the Indian economist who favors the Chinese solution but knows that he can neither order it into existence nor advise Mrs. Gandhi to implement it. I also mean the really thorny question of how special to China is the presumed success of its revolution in accomplishing the goals that Weisskopf, for example, sets out. For example, when one compares the Chinese and the Soviet experiences, one is struck by the contrast between Soviet ruthlessness and failures in its treatment of agriculture and the Chinese reasonableness and successes—a contrast that might be attributed to personal differences between Stalin and Mao or to philosophical differences between the Slavonic temperament given to extremes and the Confucian soul tempered by wisdom. But to me, the striking difference, of course, was that while Lenin, almost by accident, seized power before the revolution was accomplished in Russian society, Mao accomplished the revolution before seizing power. One has only to read Jan Myrdal's *Chinese Village* to see that the bulk of the kulaks must have been liquidated in the course of the Long War itself. Unless and until these and other such elements pertinent to the Chinese experiment's successes and failures are carefully disentangled, one has not really made much progress in effectively contrasting China with other countries, whether India or Brazil.

JOHN G. GURLEY, Stanford University: There appeared fifteen years ago, in the *American Economic Review*, an article by Wilfred Malenbaum on "India and China: Contrasts in Development Performance." As far as I know, that was the first and last time these two countries were compared in the pages of the American Economics Association's journals. This session now extends the story that Malenbaum started.

Malenbaum showed that, for the 1950's, China's growth rate of real income per capita was two to three times higher than India's. He attributed about half of the differential to China's higher rate of domestic saving and investment and the other half to the greater productivity of this investment, which he thought was the result of the Chinese putting more investment into agriculture, getting more out of their industrial investments, and centralizing the control of these capital expenditures. He concluded that: "The present analysis . . . indicates economic developments overwhelmingly favorable to the Chinese effort, both with respect to actual performance and to potential for further growth." The author, however, declared that the lesson to be drawn was not that "totalitarian methods serve better than those conceived and implemented under democracy," but rather that the Indian government simply had to get off the dime.

An interesting fact about this article is that its appearance coincided with the beginning of a sharp and rather prolonged downturn in China's economy, thereby temporarily, at least, throwing serious doubt on Malenbaum's implied forecasts. But what is even more striking about the article, from our present vantage point, is how confidently the author assumes that "development" can be measured by income per capita and can be achieved with a higher saving rate and a more efficient allocation of saving to capital formation. That's all there was to it in those innocent days!

A decade later, Gunnar Myrdal, in his monumental work, *Asian Drama*, lambasted all such circumscribed analyses, especially when they were transplanted in Asia. Myrdal criticized this type of growth theory for its simple-minded concern with only a few economic variables to the neglect of a host of other economic, social, and political considerations, all of which were closely interrelated. Developing nations, for Myrdal, were complete social systems made up of people who were engaged not only in producing, saving, and investing, but who also had certain levels of health, education, nutrition, culture, and housing, who had definite attitudes toward life and work, who acted through a variety of institutions, and who designed and implemented policies to change certain features of the system. It was Myrdal's view that "development" could only mean an upward movement of the entire social system. The main resistance to such an upward movement he saw in the retarding attitudes of peoples, reinforced by their institutions, and cemented into place with their misery and ignorance. Thus he called for a frontal attack on all the major problems at the same time. If development policies, he wrote, are "mainly directed at economic development in the narrow sense, they will prove less than effective." However, he was well aware that it was difficult "to carry out reforms that offend against established institutions and attitudes, especially since those in power . . . often share these attitudes and have a stake in preserving the institutions" they are supposed to reform. The obvious implication was that, at least in the more desperate cases, only a full-scale revolution by the peasants against the propertied classes and their official representatives could open the way for a serious development effort—though Myrdal never said this in so many words. He shied away from completing the prescription of how peasants (and workers) could organize themselves into a revolutionary force to overcome their oppressors and their own ignorance and misery; he recoiled from armed conflict.

The papers by Thomas Weisskopf and Barry Richman both contain the broader analyses urged by Myrdal, and in this respect they mark an important advance over the older formulations. Both authors, in assessing the development performances of the two countries, are immediately occupied with a swarm of social variables. Weisskopf organizes them around the three economic goals of growth, equity, and self-reliance, while Richman relates them to his more general theme about the productivity of labor. The frameworks are similar, the choice of variables to be studied about the same, and the findings of China's superiority almost identical. Despite these likenesses, there are a few differences between the two papers.

Richman's thesis is that economic development comes largely from the willingness and ability of people to work hard and effectively for inspiring goals, which in turn is a result of social development—i.e., progress in human welfare, attitudes, and institutions. Social development is mainly a matter of good ideas, of ideology and the strength of the commitment to it. Richman, therefore, suggests that China's superiority lies basically in her superior ideology (except when the commitment to it becomes too strong), which has speeded her social development, thereby raised the productivity of her people, and so boosted her growth rate of output per capita. The causal path runs from ideology to social development to labor productivity to growth rates. It follows, for Richman, that India's performance could be much improved if the Indian government would fashion and promote an ideology more conducive to development. This, I believe, starts from the wrong end.

While ideas can be a social force, they also arise from the real everyday activities of people. They come out of these activities and in turn influence them, which is to say that people, in changing the world, at the same time change themselves. Peasants who overthrow their oppressors not only change the class structure and institutions of their society but in the revolutionary process change their own conceptions of the world, their values, needs, abilities, and aspirations. The postrevolutionary economic policies flow out of this historical interaction of material and human development. Richman would presumably consider such economic policies—e.g., cooperative farming—as the materialization of a good idea, the idea of cooperation. A more fruitful way, however, is to view these policies as growing out of a revolutionary process, which itself has deep historical roots and which gives the peasants not only control over their own lives but also an understanding of the causes of their poverty and hence of its solution. This is necessarily a complicated story, involving as it does a web of interactions between theory and practice, between actual struggles and the understanding of them. The revolutionary process has unfolded in China in complex ways, while it has been developing more slowly and in different forms in India. All of this can be explained only by hard digging into the historical and dialectical development of each country's productive forces, class structure, and superstructure of values and institutions. Richman makes short work of it: China simply had a better idea. In much the same way, Malenbaum concluded that China had simply more willpower, more get up and go.

Weisskopf's answer is more carefully worked out, but it contains some similar weaknesses. China outperformed India, he tells us, because China did this and that and India did not. But why? It appears that part of the answer is of the Richman type: the greater success of the Chinese in achieving equity was because of the particular development strategy they espoused. This sounds like "better ideas" all over again. Later, Weisskopf, in accounting for China's success in attaining self-reliance, seems to revert to something like a Malenbaum-type explanation: China's better record in this area, he says, "has resulted largely from deliberate strategic choices of Chinese leadership." They simply decided to do it, while India, in principle, could have done the same thing but chose not to. Regarding growth, China's superior performance was owing to her particular approach to development—for example, to her decision to bring about a radical transformation of social and economic institutions.

These "explanations" are a bit superficial, for they suggest that "India" and "China" each sat down, like Rodin's thinker, to work out the best thing to do. China, with more brains, a firmer resolve, and wider options, won out. Weisskopf, however, unlike Richman, gets to the heart of the matter when, in his conclusion, he ascribes China's wider options to the fact that the Chinese leadership achieved "a truly revolutionary redistribution of power" and acquired "the authority and the capacity" to choose and implement its policies effectively on behalf of the peasants and workers—whereas India did not. Now that's really saying something, but it should be only the beginning, not ending, of an analysis.

Marx said that men make their own history but not exactly as they please. Richman especially, but to a lesser extent Weisskopf too, exaggerates the degree to which people freely make decisions. It would also be one-sided to say that the Indians and the Chinese did what they were compelled by history to do. But that viewpoint would be a healthy corrective to the maladies of both papers, for it would suggest that China's economic superiority over India in the past few decades is perhaps fleeting, a reflection not of more brains, more resolve, better programs, or more brilliant strategic choices, but of the two countries being in different stages of historical development. For centuries, China was a door mat for preying capitalist nations, and her record was, for long, dismal compared to Japan's. But these differences are now rapidly disappearing. India was once economically superior to Britain, then became an exploited colony. In the meantime, Britain, as the kingpin, has come and gone, and India is slowly reemerging, gathering strength, augmenting the revolutionary forces against the old powers. Uneven development has been the rule for centuries as nations have moved into and through the capitalist to the socialist mode of production. One has to have some conception of where the world is going to be able to speak wisely about present, but perhaps transitory, superiorities.

EAST-WEST TECHNOLOGICAL TRANSFER ISSUES AND PROBLEMS

International Technology Transfer: Forms, Resource Requirements, and Policies

By EDWIN MANSFIELD*

Recent years have seen a greatly increased interest in the international diffusion of new technology. For a variety of reasons, including developments in the theory of international trade and the emergence of various policy issues related to the transfer of technology, economists have been led to inquire more seriously into the process of international technology transfer. Because of the complexity of this process, and the fact that it has received relatively little attention in the past, our knowledge of it is still rather meager, particularly when compared with the demand for information by policymakers facing questions related to this process. In this paper, my assignment is to describe briefly various forms of technology transfer, to discuss the problems and costs involved in the transfer process, and to comment on some aspects of *U.S.-USSR* technology transfer.

I. Technology Transfer: Forms and Problems

In the literature concerning technology transfer, there is frequently a tendency to slur over some important distinctions, with the result that unnecessary confusion sometimes occurs. First, it is important to distinguish between *vertical* technology transfer and *horizontal* technology transfer. Vertical technology transfer occurs when information is transmitted from basic research to applied research, from applied research to development, and from development to production. Needless to say, such transfers occur in both directions, and the form of the information changes as it moves along this dimension. Horizontal technology transfer occurs when technology used in one place, organization, or context is transferred and used in another place, organization, or context. The problems involved in transferring technology from one country to another are quite different when the transfer is vertical as well as horizontal. In general, the difficulties and costs are much greater under these circumstances than if only a horizontal transfer is involved.

Second, it is important to distinguish between various types of technology. For example, as George Hall and Robert Johnson point out, it frequently is useful to differentiate between *general* technology (information common to an industry or trade), *system-specific* technology (information concerning the manufacture of a certain item or product that any manufacturer of the item or product would obtain), and *firm-specific* technology (information that is specific to a particular firm's experience and activities, but that cannot be attributed to any specific item the firm produces). Also, it is important to distinguish between several phases of the process of technology transfer. The first phase frequently is the export of a new material or product by one country to another. This phase, which Vernon Ruttan and Yujivo Hayami call *material* transfer, is often followed by *design* transfer, which is the transfer of designs, blueprints, and the ability to manufacture the new material or product in the recipient country. Finally, there is the phase of *capacity* transfer, which occurs when the capacity to adapt the new item to local conditions is transferred. Clearly, the last phase—the phase of learning how to learn as well as to use what others have learned—is quite different from the earlier phases and much more difficult and costly to achieve. It is a phase that many countries have yet to enter in the more sophisticated areas of technology.

---

* Professor of economics, University of Pennsylvania. The findings described in this paper resulted from work supported by a grant from the National Science Foundation. Also, some of them are related to the work I carried out while I was a Fellow at the Center for Advanced Study in the Behavioral Sciences at Stanford. Thanks go to Rolf Piekarz for providing relevant documents, materials, and advice concerning the material in section III.

Having touched on the differences between various types of transfers, various types of technologies, and various phases of the transfer process, I want to turn to the kinds of problems involved in the transfer of technology. Economists sometimes assume that technology is like a sheaf of blueprints and that all one has to do is ship off the right set of papers. Unfortunately, it isn't that simple or costless. For one thing, the available evidence, both recent and for earlier periods, indicates that publications and reports are a much less effective way of transferring technology than the movement of people. To transfer "know-how," much of which is not written down in any event, there is frequently no substitute for person-to-person training and assistance, some of which may have to go on for extensive periods of time.

Also, the technology that is transferred must often be adapted, if it is to meet the needs of the recipient. Because markets in the recipient country are smaller, because input prices are different, because the vendor infrastructure is different with regard to capability, cost, and quality, because of national differences in taste, climate, and so forth, the technology must be adapted. This process of adaptation can be more difficult than it may seem. Just as in the original innovation, it is very important that a proper mating occurs between the technological considerations, on the one hand, and the more purely economic considerations, on the other. (See E. Mansfield and S. Wagner.) It is easy to cite cases where, because of errors in either the private or public sectors (or both), this mating resulted in a deformed offspring.

II. Technology Transfer: Resource Costs

As emphasized in the previous section, the transfer of technology is by no means costless. Like any economic activity, it uses up resources. How great are the resource costs involved in a particular technology transfer, and what are the important factors determining these costs? As Kenneth Arrow and others have pointed out, this is an important question, and one that has received very little attention. To make any headway toward answering this question, it is necessary, of course, to construct some operational measure of the resource costs of technology transfer. When a plant is built abroad based on *U.S.* technology, there generally are outlays for engineering consultation prior to building the plant, costs of transferring engineering information concerning the process and/or product and of supervising the detailed engineering, research and development (*R&D*) costs involved in adapting the technology, and costs due to low labor productivity and poor product quality during the period when the workers are learning to utilize the new technology. David Teece, in his doctoral work (carried out under my supervision), uses the sum of these items as a measure of an important segment of the costs of technology transfer. It represents, at least crudely, the value of the resources used up in transferring to engineers and workers the relevant technology, other than that embodied in purchased equipment and materials. There may, of course, be an appreciable amount of technology transferred through the purchase of equipment and materials, but the focus here is on the costs of transferring "naked" or "unembodied" technology. Royalties are excluded (here, but not in other parts of the relevant study) because they are not a payment for resources used to transfer the technology, but a payment for the right to use it.

To estimate the magnitude of the costs cited above, data were obtained by Teece concerning twenty-six projects in which a multinational firm transferred technology abroad. In each case, a new plant utilizing the technology was built by the transferee. In the language of the previous section, most of these transfers were horizontal (not vertical as well), and practically all were system-specific and in the design phase. Table 1 shows the average and range of the relevant costs for the projects, each cost being estimated by the firm involved.[1] It also shows the mean and range of the ratio of this cost to the total cost of establishing the new plant abroad. The results indicate that these costs of transferring technology are a substantial portion of the cost of establishing a new plant abroad, particularly for the electrical and machinery firms. On the average, these technology transfer costs equaled about 36 percent of the total cost for the electrical and machinery projects, and about 10 percent of the total cost for the chemical and petroleum refining projects. (Because the technology in chemicals and petroleum refining is so different from that in machinery and electrical equipment, the sample is split into two parts, here and below.)

TABLE 1—RESOURCE COSTS OF TRANSFER OF TECHNOLOGY, 26 CASE STUDIES[a]

	Machinery and Electrical Equipment	Chemicals and Petroleum Refining
Mean Cost ($ thousands)	1,197	1,224
Range of Costs ($ thousands)	34–5,850	50–7,425
Mean Percent of Total Project Cost[b]	36	10
Range of Percentages of Total Project Cost[b]	11–59	2–22

[a] Technology transferred through the purchase of equipment or materials is excluded here. For the items included in the cost of technology transfer, see section II of the text.

[b] Total project cost equals the cost of establishing the plant, including transferring the relevant technology, obtaining equipment, and building the plant. However, for reasons given in section II, royalties are omitted.

Several hypotheses seem important in explaining the large differences among projects in the ratio of these technology transfer costs to the total cost of establishing the plant abroad. First, one would expect that this ratio would be higher if the technology that is transferred is relatively new than if it is older technology. Since newer technology tends to be less well understood (by the transferrer as well as the transferee), these costs tend to be higher. Second, one would expect, holding the age of the technology constant, that this ratio would be higher if the transferrer had built no previous plants that used this technology. The problems and costs involved in transferring an untried technology tend to be much greater than if the transferrer has been involved in many previous applications of the technology. (The former types of cases typically involve both horizontal and vertical transfers, in the language of the previous section.) Third, one might expect this ratio to be higher if relatively few firms in the world had applied the technology at the time of transfer. This variable supplements the previous one because, whereas the latter is an index of the transferrer's experience with the technology, this variable is an index of how much worldwide experience there has been with it and of how widely this experience has been diffused. (However, an identification problem may exist here because more firms may have applied the technology if this ratio is low.) Fourth, one would expect this ratio to be higher if the organization to which the technology is being transferred has had relatively little experience in manufacturing items of the general type produced by the plant. Clearly, the problems and costs involved in transferring technology tend to be greater if the transferee is relatively inexperienced.

---

[1] It is important to note two points about Table 1. (1) Some costs associated with learning by workers may be closely related to the costs of equipment and to new operating conditions. Also, the costs in Table 1 exclude costs incurred before the decision was made to build the plant. For these and other reasons, the figures in Table 1 are only crude approximations. (2) Most of these costs would have been incurred if the technology had been transferred within the United States, rather than to another country. Indeed, in some cases these costs would have been higher (according to executives and engineers) if the technology had been transferred within the United States, because of higher wage rates here. Thus, the figures in Table 1 represent the indicated costs of transferring the technology, not the extra cost of transferring it across national boundaries.

To test these hypotheses, it was assumed that

$$C_i = \alpha_o + \alpha_1 S_i + \alpha_2 A_i + \alpha_3 E_i + \alpha_4 F_i + Z_i, \tag{1}$$

where C_i is the technology transfer cost (expressed as a percentage of the total cost of establishing the plant abroad) for the ith project, S_i is a dummy variable that is 1 if there have been no previous applications of the technology by the transferring firm, 0 otherwise; A_i is the number of years that the relevant technology has been in existence when the transfer occurred; E_i is the number of years of manufacturing experience in the relevant area of the transferee at the time of the transfer; F_i is the number of firms that had applied the technology at the time of transfer; and Z_i is a random error term. Least-squares estimates of the α's were obtained, and the results seem in general to support the hypotheses listed above. Since the study on which they are based is not yet completed, these results are highly tentative. But they seem of interest since, to my knowledge, this is one of the first systematic studies of the factors associated with the resource costs of technology transfer.

III. U.S.-USSR Technology Transfer

Finally, I shall say a few words about *U.S.-USSR* technology transfer, a topic that has received considerable attention in recent years. As is well known, the Soviet Union devotes very substantial amounts of money to research and development, and has achieved notable successes in many technological areas. Nonetheless, Soviet technology lags behind that of the United States in many important areas, including computers, integrated circuits, and telecommunications. Of course, one important reason why its civilian technology lags behind is that the Soviet Union has put such high priority on military and space projects. Also, the Soviets have encountered important difficulties in getting new technology out of the laboratory and into widespread application. These difficulties seem to have been due, in part at least, to the sharp separation of *R&D* from production, which has tended to reduce the economic relevance of the *R&D* carried out and to thwart the vertical transfer of technology. Also, the incentive system facing Soviet managers seems to have discouraged innovation of many kinds, as well as the horizontal transfer of many kinds of technology.

In the past few years, there has been a marked increase in the amount of technology transfer between the United States and the Soviet Union. Although there has been widespread support for the recent spirit of detente, many observers have warned that the United States should be cautious in this area. Mindful perhaps of the Russian wheat deal, many economists have been worried about what we are getting in return for relaxation of controls. This concern has, of course, been accentuated by the asymmetry of the bargaining relationship, the State Committee on Science and Technology being in a somewhat monopsonistic position vis-à-vis American firms. Many American firms are worried about patent rights, and some have pressed for retroactive patent recognition. Further, they are worried that the Soviet market may turn out to be much smaller than some people have estimated, and they are concerned about Soviet entrance and pricing in some markets.

All in all, both the benefits and the costs to the United States of increases in technology transfer seem remarkably hazy. For example, in the view of many experts, the effects of relaxing controls on items like certain large computer systems would have only a modest effect on Soviet military capabilities (see C. Wolf). Yet others seem much more concerned about the relaxation of such controls. On the economic front, there is also a wide diversity of views, as evidenced by recent Congressional hearings on the subject. And the political or diplomatic benefits of increased technology transfer, which may be their most important rationale, are hard to predict. As Wolf remarks, "Each of us remains his own expert on the matter of political gains" (p. 7).

Turning to the much narrower topic of the transfer of economic knowledge, it is worth noting that *U.S.-USSR* technology transfer extends in a significant way to the economic aspects of science and technology, as well as to the purely technical aspects. The *U.S.-USSR* Joint Commission on Science and Technology established a *U.S.-USSR* Project Group on the economics of scientific and technological information, of which I am the *U.S.* chairman. William Baumol, Sanford Berg, Zvi Griliches, Lester Lave, Robert Willig, and I will go to Moscow this summer for the first joint symposium with the Soviets, the topic being the application of benefit-cost techniques to scientific and technological information. We hope that the proceedings will promote a better understanding in both countries of this difficult and important topic.

REFERENCES

E. Mansfield, "Technology and Technological Change," in J. Dunning, ed., *Economic Analysis and the Multinational Enterprise*, London 1974, 147–83.

——— and J. Rapoport, "The Costs of Industrial Product Innovations," *Management Sci.*, forthcoming.

——— and S. Wagner, "Organizational and Strategic Factors Associated with Probabilities of Success in Industrial R and D," *J. Bus.*, forthcoming.

C. Wolf, "U.S. Technology Exchange with the Soviet Union: A Summary Report," Rand Corp. R-1520/1-ARPA, Aug. 1974.

The Economics of East European Technology Imports from the West

By EDWARD A. HEWETT*

Governmentally imposed constraints on technology exports by Western firms to Eastern Europe[1] are no longer an important consideration outside the limited area of military hardware and related technology. Likewise, discriminatory import barriers directed particularly at East European products are also disappearing. As these political constraints lose their power, the level and composition of East-West trade will increasingly reflect more complex economic and systemic factors, which means that these trade flows are for the first time really becoming susceptible to economic analysis.

This paper is an attempt at such an analysis for East European technology imports from the West. While these are only one aspect of East-West economic relations, they are an important one both because they are closely linked to other aspects, such as East European export performance, and because they provide a particularly illuminating example of the effect of East European economic systems on the level and composition of imports from the West.

I.

"Technology" is considered here to include both specific knowledge about physical production, embodied for example in machines, designs, formulas, etc., and also more general "know-how" including, for example, organizational procedures concerning coordination in production, procurement, marketing, and research and development (*R&D*).

Eastern Europe has sought to purchase Western technology, primarily physical production technology, as part of its attempt to achieve high growth rates through continuous improvements in factor efficiency. For the small East European countries, imports are an important source of growth, technological imports an important source of technological change, and product exports the major source for financing both. Understandably, then, for most of the Western technology that East European countries buy, they seek to ensure that, in addition to increasing factor efficiency, the purchase improves the balance of payments with the West.

Through the mid–1960's the majority of East European technology imports were through direct purchase of licenses, machinery and equipment, and complete plants. However, in recent years, and particularly in the small East European countries, more complex institutional arrangements have come about which are typically grouped under the general rubric of Industrial Cooperation Agreements (*ICAs*). These are agreements between

---

* Assistant professor of economics, University of Texas at Austin. The research for this paper was conducted while I was a Visiting Scholar at the Világgazdasági Kutatóintézet in Budapest during September–December 1974. I am particularly grateful to my colleagues at the Institute, as well as to other Hungarian economists and government officials, who contributed immensely to much of the information and some of the ideas presented in this paper.

[1] "Eastern Europe" is taken here to mean Bulgaria, Czechoslovakia, East Germany, Hungary, Poland, Rumania, and the Soviet Union.

Eastern and Western firms in which usually the Western partner provides the technology embodied in machines, equipment, and licenses, and the Eastern partner repays later with products resulting from the new processes. The specific forms of the *ICAs* vary. (United Nations, 1973, provides a general review.) They can involve, for example, simple purchases of complete plants or licenses in exchange for later payment with the resulting products; subcontracting, in which the Western partner provides technology and possibly some inputs for producing a certain component and the Eastern partner repays in a fixed amount of components; coproduction and comarketing, where the partners specialize in parts of a single production process, assemble the final product in one or both countries, then divide up marketing areas; and finally, joint operations on third markets. One other form of *ICA*, the joint venture, is only possible in Hungary, Rumania, and Yugoslavia, and only in Yugoslavia (which is not being considered in this paper) has it assumed any real significance. Joint venture agreements allow for up to 49 percent ownership, but there are substantial differences among the three countries in most of the major provisions for the agreements. (C. McMillan and D. St. Charles provide a good comparison.)

While some *ICAs* do involve technology transfer from the East to the West, the flow is predominately in the other direction. East European economists argue that the *ICA* is the most efficient form for importing Western technology from several points of view (K. Szénásiné-Zboróvári, M. Simai). First, the *ICA* seems to increase the probability that the East European country can obtain the most modern technology for sale in the West and that it will be assured of adequate help from the Western partner in absorbing the technology, since the Western firm must accept and use, or market, the resulting products. Second, the Western firm will force the Eastern enterprise to meet higher quality requirements and guarantee smoother operation of the production line than are typical in most socialist enterprises. Third, *ICAs* are expected to improve access to Western markets, both because Eastern enterprises can learn from the Western firm's marketing organization and because they can avoid some of the high costs attached to independent entry into Western markets. Finally, the so-called "complex-barter" aspects of *ICAs* mean that little money is involved and that there is a more or less guaranteed balance of payments effect arising from a fixed export commitment signed when the technology is imported (United Nations 1970).

Even accepting these propositions as being substantially correct, they only address one side of the question, that of the most efficient form for extracting from Western firms the technology which Eastern Europe wants to acquire. The other side is whether Eastern Europe, given its goals and constraints, is buying the correct amount, quality, and quantity of Western technology, and if it is using it as efficiently as possible.

Very little useful information is known about the actual performance of *ICAs* in that regard, in part because they have only recently begun to grow beyond the insignificant level of the early 1960's. A more fundamental impediment to learning more about *ICAs* is that most of the requisite information resides in Western and Eastern enterprises which will not allow the detailed access to their records and organizations necessary for serious research.

In an attempt to gather the information which does exist and to derive hypotheses useful in more detailed research, I have undertaken a study of the Hungarian experience with *ICAs*. Hungary now accounts for about one-half of all *ICAs* existing between Eastern Europe and the West (interview material); thus it provides an excellent case study of East European problems with *ICAs*. Although it might appear that Hungary is a special case because of the 1968 Economic Reform, in fact that Reform provides crucial new information about enterprise attitudes toward technology transfer in Eastern Europe. It must be emphasized that the conclusions which follow are very tentative; and firmer conclusions will not be possible until, and if, more data are published and detailed case studies are completed.

II.

Only a general outline of the post-1968 Hungarian economic system is necessary here (for details see B. Balassa, D. Granick, and M. Tardos). Hungarian enterprises receive no obligatory plans from the government; they alone are legally responsible for plans and contracts covering current outputs and inputs, investments, and foreign trade. The government still constructs short-, medium-, and long-term plans; however, they are much less detailed than in the old system, containing projections for aggregate and sectoral variables and considerable detail on the economic and administrative instruments which will be used to influence enterprise decisions.

Many enterprises have monopolies for the products they produce; no important products are produced or sold in anything approaching a competitive atmosphere. High tariffs, a strict import-licensing system, a foreign exchange monopoly, and export subsidies combine to protect enterprises from competition. There are strong price controls for many products, particularly consumer goods. Ministries retain substantially more influence over the enterprises than the market, primarily as a legacy of their absolute power in the old system, but also through their legal rights to appoint and dismiss managers, to control investment which goes through the state budget, and through numerous "favors" they can do for enterprises. (See Z. Hegedüs and Tardos for an interesting case study of ministry-enterprise relations.)

Consistent with the Reform, practically all proposals for *ICAs* originate with enterprises rather than the Hungarian government; in fact, the majority of those actually originate with Western firms' proposals to their Hungarian counterparts. Government policies serve to encourage *ICAs* through preferential credits for investments which will increase net exports and through tariff concessions on imports coming through *ICAs*. The Hungarian National Bank and the Foreign Trade Bank administer the credits, while the Interministerial Cooperation Committee (Tarcaközi Kooperációs Bizottság) decides on the preferences. This Committee has a strong veto power over proposed *ICAs*, which it apparently exercises mainly to prevent *ICAs* in industries where the government has decided that imports will be the sole source of domestic supplies (or at least increases in domestic supplies).

Almost all of the more than 300 *ICAs* in force now in Hungary are rather small subcontracting or coproduction arrangements. The Eastern partner is usually producing the least capital-intensive parts of the products (or the least capital-intensive finished products) using substantial amounts of imported Western materials and possibly some imported capital, but mostly its own excess capacity; then, either all or most of the product goes back to the Western partner as an input or to be marketed. For the Western partner, these "small deals" are quite profitable, since they shift the more labor-intensive, less profitable processes to an area where semiskilled and skilled labor is relatively cheap and a good deal more dependable (no strikes). There are other motivations for Western enterprises, such as avoiding high tariff and nontariff barriers to direct imports from the West. But the main concern here is what the motivation is on the Eastern side.

That is difficult to ascertain and is currently the subject of debate in Hungary. First of all, it is interesting, and quite disturbing to the Hungarian government, that large Hungarian enterprises are the *least* interested in *ICAs* of any kind; practically all of the *ICAs* involve small and medium-size Hungarian enterprises. Even for those latter enterprises the motivation to enter into *ICAs* is probably not to buy new technology or to transform the technological level of their enterprise; the agreements themselves reveal contrary preferences. In fact many *ICAs* are probably traditional trade under the name of "cooperation," through which firms can acquire otherwise unobtainable loans, investment equipment, and financial advantages (K. Zboróvári, p. 439). In sum, there is now a contradiction between the Hungarian government's ambitious aspirations and Hungarian enterprises' much more modest goals concerning technology transfer.

The source of this contradiction would seem to be closely linked with the fate of the 1968 Reforms. One of the disappointments of those Reforms is that, despite a clearly specified goal to increase competition in the Hungarian economy, very little actual progress has been made in that direction. As stated earlier, most enterprises enjoy monopolies in the products they produce. Their primary constituent, particularly if they are large enterprises, is the ministry rather than the market. The enterprises' organizational structure remains as highly centralized as it was before the 1968 Reforms, a frozen memory of the previous economywide hierarchy (M. Timár). Their activity is almost totally devoted to production, to the exclusion of marketing or research and development.

These enterprises are quite truncated by Western standards, which explains their difficulty in competing on Western markets. The most successful among Western enterprises are characterized by quite sophisticated organizational structures which execute and coordinate marketing, research and development, and production. (P. Lawrence and J. Lorsch). This explains why, in the *ICAs*, the Hungarian enterprise typically is involved only in production while the Western enterprise is involved in research and development and marketing, as well as production. The Hungarian enterprise usually appears to be most efficient in production and relatively much less efficient in *R&D* and marketing. By following the dictates of static comparative advantage, the *ICA* serves to perpetuate some of the major weaknesses of the Eastern enterprise, rather than eliminating them.

The organizational changes which would make Hungarian enterprises independently competitive on Western markets would be quite difficult and costly, requiring a substantial redistribution of power in the enterprises from directors to lower functional units, the formation of new units, promotion of younger people, etc. It seems highly unlikely that Hungarian enterprise directors will voluntarily introduce such extensive changes, even given a great deal of pressure on the part of the government. Much stronger forces than government directives will be necessary.

Probably the only truly effective mechanism for compelling these changes is increased competition in Hungarian markets, primarily through lowered barriers to foreign trade, and a substantial curtailment of the extreme power of ministries over enterprises. While there are proposals to that effect (e.g., Tardos), it seems unlikely that such drastic changes will occur soon. In the first place, it would be very expensive to introduce competition rapidly and extensively into the Hungarian economy since it would involve scrapping a significant portion of the capital stock, many bankruptcies, and a substantial redistribution of the labor force. Politically, that would be difficult to handle, and from a financial point of view would require an extremely large net capital inflow into Hungary. Also, the enterprises, the ministries, and the labor unions in Hungary comprise a formidable protectionist lobby which effectively opposes proposals for substantially opening the economy to foreign competition.

This means that the contradiction between the Hungarian government and Hungarian enterprises is likely to remain in the near future; *ICAs* will probably not soon serve as a vehicle for substantial technological transformation of Hungarian industry. Hungarian enterprises are not really interested in being "transformed"; their market situations are far too comfortable for that. On the other hand, the government will surely continue its pressure for the real technical and organizational transformation so necessary to improved factor efficiency and export performance on Western markets, as a natural continuation of its first attempt at comprehensive reform in 1968.

III.

Thorough studies of the Hungarian experience with *ICAs* are necessary to confirm the rather preliminary hypotheses presented here. Similar studies for other socialist countries are also necessary to see to what extent the Hungarian experience is unique. Most of what has been said here probably applies to other East European countries, including the Soviet Union. The overcentralization of enterprise organizations seems a common characteristic of all East European countries, and so must be the impediments that overcentralization creates for the importation and use of Western technology.

Previous to the introduction of the Hungarian Reform, the tendency was to blame problems in industrial cooperation, and East-West trade in general, on the fact that enterprise initiative was stifled by overcentralized control of their activities. And indeed there is substantial truth in that proposition. Between 1963 and 1967 Hungarian enterprises concluded about thirty *ICAs*; then in the following three and one-half years, when enterprises and not the government were the source of the proposals, 137 agreements were concluded (Zboróvári, p. 15). Probably about another 150 more have been concluded through mid-1974 (interview material). Still, it is obvious that enterprises in Hungary exhibit a quite conservative form of initiative and that centralization within the enterprise is yet another tenacious barrier to realizing the full potential for *ICAs* and thus for technology transfer.

REFERENCES

B. Balassa, "The Economic Reform in Hungary," *Economica*, Feb. 1970, *38*, 1–22.

D. Granick, "The Hungarian Economic Reform," *World Politics*, Apr. 1973, *25*.

Z. Hegedüs and M. Tardos, "A vállalati vezetök helyzetének és motivációjának néhány problémája" ("Enterprise Directors' Situation and Motivation, a Few Problems"), *Közgazdasági Szemle*, Feb. 1974, *21*, 162–73.

P. Lawrence and J. Lorsch, *Organization and Environment: Managing Differentiation and Integration*, Boston 1967.

C. McMillan and D. St. Charles, *Joint Ventures in Eastern Europe: A Three-Country Comparison*, Ottawa 1973.

M. Simai, "International Interfirm Cooperation in East-West Relations," *Marketing in*

Hungary, Jan.–Mar. 1973, 7–11.

K. Szénásiné-Zboróvári, "A tökes kooperációk néhány problémája" ("Capitalist Cooperation, Several Problems"), *Külgazdaság*, 1973, 430–41.

M. Tardos, "The Problems of Economic Competition in Hungary," mimeo., Budapest 1972.

M. Timár, "Results of the New System of Economic Control and Management and its Further Development," *Acta Oeconomica*, 3–4 1973, *10*, 277–301.

K. Zboróvári, "A nemzetközi termelési kooperáció néhány elméleti és gyakorlati kérdése" ("International Production Cooperation, Several Theoretical and Practical Questions"), *Vezetöképzés*, 4, 1972, *3*, 10–16.

United Nations Economic Commission for Europe, "A Research Note on Industrial Cooperation as a Factor in the Growth of East-West European Trade," *Econ. Bull. for Europe*, 1, 1970, *21*, 67–88.

———, "Analytical Report on Industrial Cooperation Among ECE Countries," E/ECE/844; 14 Mar. 1973.

Problems of Technology Absorption in Chinese Industry

By Thomas G. Rawski*

The ability to respond promptly and effectively to demand changes is one important distinction between advanced and backward economies. Except for small trading nations, such flexibility presupposes a highly developed capital goods sector which can meet current investment requirements. What factors affect the capacity of capital goods producers in developing countries to emulate their more advanced counterparts by transforming domestic production possibilities in response to shifting economic patterns?

The experience of China, a nation which has achieved important gains in expanding the flexibility as well as the size of domestic capital goods output, offers a useful framework for investigating this question. China's capital goods industry is presently dominated by two quite distinct groups of firms: recipients of major investment outlays and Soviet equipment shipments during the 1950's, including some with pre-1949 histories of Japanese ownership or management; and older, smaller enterprises which constitute the legacy of prewar industrial advance in the private sector and have typically received only minor infusions of imported equipment and technical aid since 1949.

The contrasting performance of these groups during the 1960's, a decade in which economic difficulties and international isolation forced domestic producers into ambitious attempts to improve quality, variety, and manufacturing techniques, provides a test of the relative importance of modern equipment and of industrial experience in determining innovative capacity in capital goods production.

With industry performance now judged primarily by the contribution of current output to immediate economic needs rather than, as in the 1950's, by its sheer volume, the second group of firms has established a clear position of technological leadership, most notably in the machinery sector.

The success of older units in substituting experience for investment points us toward models which emphasize the accumulation of skills as well as fixed assets in enlarging society's production possibilities, an approach emphasized by historians of American and Japanese development. The contributions of long-established enterprises illustrate the importance of prewar advances even in sectors which have recorded the largest increases in output and capital stock since 1949. And the inability of newer, Soviet-aided units to match the innovative performance of less well equipped plants suggests that, particularly in the presence of uncertainty, the benefits of international and domestic division of labor and specialization in capital goods production must be weighed against the less obvious gains which arise from widespread participation in the task of designing, manufacturing, and repairing new types of equipment.

---

* Assistant professor, Department of Political Economy, University of Toronto. E. S. Rawski commented on an earlier draft. Notes and references are available from the author.

I.

Suppose that a task T requires skill levels S_{1T}, S_{2T}, . . . , S_{nT} in each of n areas. A simple extension of the concept of "learning by doing" is the hypothesis that repeating T not only lowers its cost, but also raises potential in the various skill areas to new levels P_1, P_2, . . . , P_n which exceed the initial minima required to complete T. The resulting opportunity to develop previously unattainable skill combinations will encourage entrepreneurs to attempt more difficult tasks T^* if they believe that $P_i \geq S_{iT^*}$ for $i=1, 2, \ldots, n$ and that the gains from innovation outweigh the cost of experimentation and the risk of failure.

Although industry's skill frontier is influenced by the amount and quality of equipment, materials, education, and management, China's recent experience shows that during at least some phases of industrialization production experience may be a key determinant of the level and fungibility of industrial skills.

The significance of skill accumulation can be seen from the development of engineering both before and after World War II. China's prewar engineering industry emerged from a cycle of backward linkage and import substitution associated with new modes of production and transport imported from Europe. The appearance of steamships and railways and the subsequent growth of consumer manufactures created demands for repair services and spare parts which were initially supplied by foreigners. As Chinese experience and capital grew, native mechanics replaced foreigners, and import substitution then spread into commodities, with Chinese firms acquiring significant market shares first for parts and later for a wide range of equipment as well.

This pattern of evolutionary growth is illustrated by Shanghai's Talung Machinery Works, which began in 1902 as a ship repair specialist and, taking full advantage of the absence of scale economies and the transferability of basic engineering operations, moved successively into textile machine repairs, parts manufacture, and, finally, production of complete sets of cotton-spinning equipment.

The skills acquired by veterans of Talung and other private firms and by men who studied abroad or worked in railway shops or foreign-owned factories supported rapid growth and diversification during the final prewar decade (1927–37). In Shanghai the number of metal-processing and machinery producers using mechanical power and employing over 30 persons rose from 82 to 173 during the depression years 1931–33 alone, and output, capital, and employment for all firms in these sectors rose by about 300, 100, and 60 percent. Similar growth occurred on a smaller scale in Tientsin, Wusih, Chinan, and other cities.

The ease with which firms manned by experienced Chinese could make the transition from repair work to manufacture is reflected in the success of domestic firms in penetrating a growing range of machinery markets, particularly at the lower end of the price-quality spectrum. One survey lists 99 manufacturers of textile machines: spindles, looms, stocking frames, bowing machines, equipment for reeling and weaving silk, cotton gins and dyeing equipment. Manufacturers of milled rice, flour, cooking oil, canned goods, books, ink, leather, matches, cigarettes, candles, and soap also purchased domestic equipment.

Private engineering firms also supplied investment goods to the producer sector. In addition to producing equipment for self-expansion, machine shops sold machine tools to customers in China and in Japanese-controlled Manchuria and turned out a variety of equipment for chemical plants, cement works, smelters, mines, and railways.

By the 1930's, the engineering sector, although still small, could already help enlarge the modern sector. During World War II, domestic producers reequipped battle-damaged factories and manufactured munitions in both occupied and free China.

Despite these gains, the economic plans implemented by the new Communist government during its first decade (1949–60) made scant provision for modernizing and enlarging the inherited private engineering sector. Instead, the first Five-Year Plan (1953–57) assigned nearly half of overall industrial investment to 145 Soviet–aided projects. With few exceptions, these projects either created new facilities, such as the engineering complex at Loyang, or expanded Japanese-built plants, such as Anshan's steelworks, in which earlier Chinese participation had been limited to supplying semiskilled labor.

Shanghai, the center of China's inherited industry, turned out 22 percent of machinery output and 19 percent of national industrial output in 1957, but received only three of the Soviet-aided projects and a mere 2.5 percent of overall investment outlays during 1953–57.[1] Tientsin fared no better, gaining only a 7 percent increase in its machine tool stock during 1953–57, years in which gross fixed investment in industry amounted to over double the 1952 capital stock.

Security considerations and a desire to reduce the influence of the urban bourgeoisie may have justified this policy during the early 1950's, but the failure of Shanghai's investment share to rise above 4.5 percent of the national total even after the Korean truce, the 1956 socialization campaign, and widespread advocacy of coastal industrial expansion shows that planners expected little of China's older and smaller industrial units.

China's Second Five-Year Plan (1958–62), compiled with the expectation of continued large-scale acquisition of Soviet blueprints, equipment, and technology, was disrupted by the Great Leap Forward (1958–60) and shattered by the sudden withdrawal of Soviet technicians in 1960 and an agricultural depression caused by successive harvest failures in 1959–61.

The ensuing crisis led to major unanticipated demand shifts. Peking's belated decision to divert more resources to the farm sector increased requirements for fertilizer, insecticide, tractors, pumps, and other agricultural producer goods. Growing Sino-Soviet animosity disrupted Soviet supplies of petroleum, military goods, and equipment needed to complete unfinished construction projects.

With exports falling, grain imports rising, and the industrial powers unwilling to provide needed capital goods, the task of altering domestic production functions to meet the new situation fell squarely upon China's machinery producers.

The considerable success with which engineering firms met the challenges of the 1960's is largely attributable to the achievements of enterprises and regions which had gained little from the investment boom of the 1950's. The petroleum sector, which has achieved a tenfold output increase since 1960, is equipped primarily by converted engineering plants including former textile machine repair works. Descriptions of the petroleum machinery industry make specific reference only to smaller and older plants: Shanghai's complex, including the Talung Machinery Works mentioned above, and the venerable Lanchou General Máchine Works, a provincially managed descendant of a nineteenth century arsenal.

---

[1] Data on Shanghai investment shares are taken from an unpublished study by Nicholas R. Lardy of the University of Michigan.

The role of older firms is even clearer in the case of chemical fertilizer, a sector in which output rose from a possibly exaggerated Great Leap peak of 2.5 million tons to 9 million tons in 1965 and 20 million tons in 1972, again without major equipment imports. Domestic fertilizer equipment first appeared in quantity during 1962; Shanghai led the way, supplying 25,000 tons of equipment to one major plant. While expanding the volume of output, Shanghai firms also expanded their product line, turning out China's first equipment for urea manufacture in 1965. Small fertilizer plants, which have supplied over 80 percent of incremental output since 1965, obtain equipment from many sources, but the core of the equipment industry remains in Shanghai, where a group of over 400 factories supplied 300 sets of equipment for synthetic ammonia plants between 1970 and 1972.

Older units take the lead in introducing such new techniques as "electronic spark processing, forge pressing, vacuum casting, iso-ion cutting, argon-arc welding, electroslag welding, hot riveting, chemical and unlined hearths" into Chinese industry. Talung Machinery Works, which has pioneered in manufacturing petroleum equipment, compressors, and machinery for producing artificial diamonds, illustrates the continuing innovative success of older engineering firms.

Why are older and smaller firms better suited to developing new products and techniques than larger units? Capital cannot provide an explanation, for we know that investment planning bypassed the inherited engineering sector until 1960 at the earliest. In Shanghai's case, municipal reports indicate low or declining levels of capital formation until 1965.

The advantage of small units may arise from diseconomies of scale. Chinese observers report that smaller firms are "more compact and . . . more flexible in adjusting their production." But at Shenyang's #1 Machine Tool Works, a Soviet-built enterprise employing over 5,000 workers, organization is "swollen and ineffectual; technical control . . . highly complicated . . . the potentialities of the equipment and personnel cannot be fully developed and utilized and production technology is not likely to advance and improve."

The great strength of Shanghai and Tientsin lies in experience. With their "skilled veteran workers and experienced technical persons" and superior development of interenterprise cooperation, another legacy of the past, "old industrial bases and old enterprises . . . find it easier to tackle . . . complicated technical problems than new enterprises and new industrial bases."

With these advantages, it is the established centers which are best able to copy and modify foreign equipment samples, to extract useful information from foreign technological publications, and to apply it to current domestic problem areas. More generally, experienced firms act as technological intermediaries between the mass of Chinese producers, whose mission is to attain "advanced national levels" of quality, cost, and technique, and the outside world, whose standards become the target of Shanghai's technological aspirations.

Technical advance penetrates the economy along well-developed lines of communication which include worker exchanges among units of differing technical levels, national and local meetings focused on particular problems or on "exchanging advanced experiences," national journals for machinery, chemicals, metallurgy, and other industries, and several Shanghai journals devoted to publicizing technological gains of Chinese and foreign origin.

Campaigns to "emulate the advanced" and to "learn from Shanghai" illustrate this diffusion process. As one headline put it, "The Nation Looks to Shanghai for Production Experiences." Thousands of workers and managers, including groups from Peking, Wuhan, and other large industrial centers, visit Shanghai to "learn new techniques and to exchange knowledge" and to see that since "conditions at . . . Shanghai . . . were not any better" than elsewhere, massive investment is not a prerequisite for technical advance.

Manufacturers of steel, machine tools, diesels, pumps, pneumatic tools, sewing machines, and many other commodities judge their products by comparing them with Shanghai's, and to a lesser extent with Tientsin's. Even in cable manufacture, a field dominated by the Soviet-aided Shenyang Cable Works, "research units were established in Shanghai as centers of research and design and the experimental base for the technology of the cable industry."

During the past decade, an isolated China has taken significant steps toward overcoming economic difficulties of a magnitude which other large developing nations have rarely faced without extensive foreign assistance. These gains demonstrate that, despite the obvious limits to technical sophistication in all branches of industry, China has developed an effective mechanism for implanting foreign innovations into Chinese industry, for spreading new techniques and products from advanced to backward enterprises and regions, and, in some areas, for devising solutions to uniquely Chinese technological problems.

II.

What can these results contribute to our knowledge of the problem of import substitution and technology transfer shared by all developing economies?

Foreign capital goods typically offer the quickest and cheapest method of injecting new activities into a backward economy. China, for example, could not have begun mass production of tractors, trucks, ball bearings, radios, and a host of other products during the 1950's without massive equipment imports.

Reliance on imported investment goods also has its disadvantages. China is only one of many importing nations which have found it difficult to match local skills, materials, and scarcity relations to the requirements of equipment designed for a foreign economic milieu. These problems often erase at least part of the anticipated advantages of imported equipment and techniques.

More serious is the tendency of a trade-oriented industrialization strategy to curtail the skill formation process which accompanies the search for domestic solutions to technical difficulties. Growth of skill and of the resulting capacity to adapt production to current needs is especially valuable in engineering, the one industry with the potential of supplying all sectors with the means to lower costs and advance productive methods.

Unless domestic engineering producers can begin to display the flexibility which is the hallmark of their advanced-country counterparts, a developing economy runs the risk of finding itself unable to cope with demand shifts too large to be resolved by foreign trade alone. In China's case, only the unanticipated flowering of innovative capacity among older machinery producers prevented such a debacle during the 1960's.

The conclusion to be drawn from this history is that, at some points in the industrialization process, the long-term impact of capital goods imports, which can and do raise the short-term rate of embodied technical progress, may be negative. This possibility arises because imports prevent domestic firms from learning to implement new techniques without purchasing foreign equipment. If exchange stringency or political problems disrupt foreign equipment supplies, domestic capacity to absorb new methods suddenly becomes a constraint on the rate of embodied technical change throughout the economy.

This view, with its emphasis on the long-term benefits of skill accumulation, seems to lie behind the broad Chinese consensus favoring a high degree of national and regional self-reliance in the manufacture of machinery and other producer goods. While accepting imports as a means of breaking technological bottlenecks—recent contracts to purchase eight large-scale fertilizer plants provide an example—the Chinese have kept engineering imports well below 10 percent of domestic output and will probably continue to do so.

Internally, the sequence of repair, parts manufacture, and full production is encouraged in both urban and rural areas. Former repair shops have begun to manufacture trucks, locomotives, and agricultural equipment. Small rural producers of fertilizer, cement, iron, power, coal, and machinery are expected to repair and often even to manufacture their own equipment.

The inevitably mixed results of local self-reliance are tolerated because the Chinese believe that widespread industrial experience will in time reproduce the development sequences which are responsible for the present innovative capacity of older industrial centers. Thus reports of success, as in the thriving factories of Kwangtung's prosperous Chungshan county, alternate with instances of failure, as in Hopei's Kaoch'eng county, where efforts to combat drought foundered because locally produced engines turned out to be useless.

It is difficult to gauge the extent to which Chinese experience and policy in international and internal technology transfer are relevant to the needs of other industrial latecomers. Many nations lack the industrial foundation inherited by the People's Republic; elsewhere, extensive import substitution has failed to stimulate domestic innovative capabilities.

Furthermore, the process by which skill formation extends society's production frontier is both discontinuous and hard to observe, and the resulting innovations can easily elude the economist's productivity measures. Despite these difficulties, the contribution of cumulative industrial experience to China's recent economic gains underlines the desirability of further research into the skill dimensions of industrialization.

TAXATION OF NATURAL RESOURCES

Incidence of World Taxes on Natural Resources with Special Reference to Bauxite

By MALCOLM GILLIS AND CHARLES E. MCLURE, JR.*

Recent increases in taxes on natural resources have focused worldwide attention on the question of their incidence. Who, indeed, does pay these taxes? In the first section we describe a conceptual framework for ordering analysis of the question based on the comparative statics reasoning of tax exporting and general equilibrium incidence analysis. (See McLure, 1964, 1967, and forthcoming.) Section II applies this framework to one concrete case, the heavy production tax Jamaica imposed upon bauxite last summer. Little attempt is made in either section to analyze the fascinating game-theoretic and other dynamic aspects of the mutual adjustment process set off by taxes on natural resources.

I. The Conceptual Framework

The general determinants of the incidence of taxes on natural resources can be presented as in Table 1, though the compartments are not watertight and many nuances are ignored.

The first consideration is whether taxing nations dominate the relevant market for the resource. The relevance of this point is twofold. First, nondominance implies a relatively elastic demand curve for the taxing nation's output, regardless of the industry's demand curve. Second, taxes can be employed to tap resource rents. But if the tax levied by a nondominant country exceeds differential rents resulting from differences in quality and accessability of ore and transportation costs, further exploitation may become uneconomical. (See H. Hughes.) We ignore the possibility of coordinated action by consumer nations.

A second consideration is whether capital is sufficiently mobile to flow from the taxed industry in response to potential reductions in earnings. If capital is mobile, its supply is probably elastic, as seen by the extractive industry in the taxing country (countries). If not, capital—and the accompanying technology—earns essentially quasi-rents. Whether capital is "mobile" depends upon whether we are considering the short run or the long run of economic theory, or, in concrete terms, upon the time phasing of depreciation and maintenance of productive facilities and the level of gross investment in the industry (and, in some instances, upon the nation's exchange control policies). Finally, though we suppress it in Table 1, immobility of labor has implications analogous to those of immobility of capital. But the elasticity of supply of labor depends more crucially upon the possibility of absorbing labor elsewhere in the economy, that is, upon the elasticities of factor sub-

---

* Institute fellow, Harvard Institute for International Development, and lecturer in economics, Harvard University, and professor of economics, Rice University, respectively. We wish to acknowledge useful comments Peter Mieszkowski, Michael Roemer, Joseph Stern, and Louis T. Wells made on earlier drafts of this paper, but to hold them blameless for errors of analysis or interpretation.

TABLE 1—LIKELY INCIDENCE OF TAXES ON NATURAL RESOURCES[a]

I. *Market Dominance* by Taxing Nations	II. *Nondominance* by Taxing Nations
A. Capital Mobile	A. Capital Mobile
Case	Case
1. Competition	7. Competition: resource rents
a. Profits tax: resource rents	8. Monopoly: resource rents
b. Production or export tax: consumers	9. Oligopoly: similar to monopoly case
2. Monopoly:	
a. Profits tax: monopoly rents and resource rents	
b. Production or export tax: one-half by consumers; one-half by monopoly and resource rents	
3. Oligopoly: intermediate between competition and monopoly, depending upon prior success in maximizing joint profits	
B. Capital Immobile	B. Capital Immobile
4. Competition: quasi-rents and resource rents	10. Competition: quasi-rents and resource rents
5. Monopoly: quasi-rents, monopoly rents, and resource rents	11. Monopoly: quasi-rents, resource rents, and monopoly rents
6. Oligopoly: similar to monopoly case, unless joint profits are not maximized. (Shifting to consumers may be possible.)	12. Oligopoly: similar to monopoly case

[a] Applies only to noncreditable portion of corporation income tax. Credited portion borne by crediting nation. Labor receiving higher than competitive wages may share in burdens borne by monopoly and resource rents. Constant costs are assumed to prevail if capital is mobile.

stitution and relative factor intensities in the taxed industry and the rest of the economy.

The third consideration is the industrial structure in the production and distribution of the taxed resource. For simplicity we concentrate upon three traditional descriptions of market structure: competition, monopoly, and oligopoly. Unless the industry is perfectly competitive, there will be monopoly (or monopsony) rents that may be affected by the tax. But this classification does not do justice to the richness of industrial structure observed in individual cases. In particular, the "oligopolistic" characterization tells us little except that there are few firms. In any concrete case we must attempt to determine more precisely how the firms behave, including whether nor not they collude with taxing countries. Thus our conceptual framework is useful in structuring the problem, but is no substitute for detailed analysis.

A final consideration omitted from Table 1 is the prevalence of labor rents. Unionization and industrial concentration often go together, and whether unionization occurs or not, in many cases wages are higher in the taxed industry than in the rest of the economy because of the tendency to equalize, at least partially, wages paid to nationals and expatriates. In either event, local labor shares in resource and monopoly rents.

We can identify twelve cases potentially worth examination, or twenty-four, if we allow for labor rents. Other factors are somewhat different in kind and do not readily lend themselves to taxonomy. First, the incidence of corporation income taxes is more complicated than that of export duties and production taxes. It involves taxation of the return to equity capital as well as economic profits, transfer prices within multinational firms, and, perhaps more importantly, foreign tax credits sometimes available only for income taxes. In general, so long as the taxes in the producing nations do not exceed allowable credits, the credit, made at the expense of taxpayers in the nation granting it, effectively reduces the amount of corporation income tax borne by consumers and recipients of factor incomes.

A second consideration involves whether the contracts typical of the industry can be adjusted to reflect changed fiscal conditions. For our purpose, the "owner" is the party who has an economic interest in the resource over the time period in question, as determined by contractual arrangements, though in many cases the reference may not be legally accurate. That is, he is the one obtaining "resource rents."

If taxing nations dominate the market, other cost and demand conditions become relevant. Substitutes increase the industry's elasticity of demand. Being unimportant—the fraction the resource cost is of total costs—is another crucial determinant of the elasticity of demand. Marginal cost rises more rapidly the less mobile are factors. Beyond that, the quality of the resource and the ease of extracting, processing, and transporting it determine differential resource rents.

The final consideration is a purely mechanical matter. Once we know who—in terms of factor shares and consumer groups—bears the tax, we must allocate the tax to their nations of residence.

Table 1 summarizes the likely incidence of taxes in each of twelve cases. For convenience we assume that if capital is mobile the taxed resource can be exploited at essentially constant costs and that labor outside the taxed sector generally would not share in the tax burden through tax-induced inflows of labor from the taxed sector. Moreover, we suppress repeated mention that labor might share in any reduction of rents. Finally, we distinguish between owners of natural resources (recipients of resource rents) and capitalists exploiting the resources, though in practice the two groups may be virtually identical. Capitalists earn returns to capital (normal returns in the long run; quasi-rents in the short run) and perhaps monopoly profits.

Case 1: Competitive firms subject to production or export taxes in nations dominating the world market would shift the taxes in large part to consumers (except as restrained by substitutes) if capital is mobile. Income taxes would be borne largely by recipients of resource rents, and perhaps by consumers. Its mobility would allow capital to escape the tax.

Case 2: In the constant-cost case a monopolist could shift only one-half a tax on production or export to consumers, the rest being absorbed through reduced monopoly and resource rents. The profits tax would come largely from monopoly and resource rents.

Case 3: The results for oligopoly would resemble those for either competition or monopoly, depending upon how well the firms had done in maximizing joint profits.

Cases 4–6: With capital immobile, either a profits tax or a tax on production or export would generally be borne by recipients of resource rents and quasi-rents of capital and, where relevant, monopoly rents. In these cases, as others, reductions in rents may be borne in part by resident labor, but would ordinarily result in substantial tax exporting. A qualification to this general conclusion is that oligopolistic firms may succeed in shifting part of the tax to consumers if they have not previously maximized joint profits.

Case 7. In this case both the demand for the resource and the supply of capital are perfectly elastic. Thus a tax can neither be shifted to foreign consumers nor reduce the normal earnings of capital. Either a profit or an export or production duty must be borne by owners of the natural resource.

Case 8: The marginal revenue of a monopolist dominating the world market, to which marginal cost in the taxing nation is equated, is determined in the world market. And, by assumption, so is the return to capital. Thus, again, no tax would be shifted forward in higher prices or reduce the return to capital per se. Rather, the burden of a profit, production, or export tax would be borne by owners of natural resources.

Case 9: Again, the results for oligopoly should resemble those for competition and monopoly, namely, burden on resource rents, as in cases 7 and 8.

Cases 10-12: Demand is perfectly elastic, but capital is fixed in supply and therefore earning quasi-rents (and monopoly rents, in cases 11 and 12). Either kind of tax would be borne largely by recipients of various rents.

In summary, then: we expect either profits taxes (in excess of credits) or commodity taxes to be borne largely by recipients of various rents (quasi, resource, and monopoly), including labor, except for the long-run case of a production or export tax levied by nations dominating world markets. If natural resource industries subject to commodity taxes by dominant nations were competitive and substitutes were unimportant, virtually the entire tax might be shifted to consumers. But if monopoly or joint profit maximization by oligopolistic firms prevailed, only about half the tax would be shifted to consumers.

In any event, most countries in which exploitation of natural resources is under foreign ownership could probably export the lion's share of taxes on the resources. First, substantial amounts of creditable taxes are paid by foreign treasuries. Second, if taxing nations dominate the world market, half or more of commodity taxes can be exported to consumers, especially if capital is mobile. Third, reducing the quasi-rents accruing to immobile capital involves substantial tax exporting, at least in the short run. Fourth, monopoly and monopsony rents also accrue largely to nonresidents. Finally, foreigners often effectively own natural resources and would be burdened by any tax that reduces resource rents. The primary caveats involve the sharing of local labor in the reduction of rents and the risk of going so far as to deter future gross investment in the industry, or even in the country.

II. The Bauxite Case

In mid-May Jamaica announced an increase in its bauxite taxes of *U.S.* $170 million per year and the intention to increase Jamaican participation in the ownership of bauxite and alumina operations. (See M. Manley. Bauxite is the principal ore from which aluminum oxide—alumina—is refined for later smelting into aluminum ingot. Total tax receipts in 1973–74 were *U.S.* $357 million; see Jamaica, Ministry of Finance.) Due to the difficulty of determining an arms-length price for bauxite and profits attributable to operations in Jamaica, the tax was levied on production, rather than on profits, and was geared to the price of aluminum ingot. Moreover, it included a presumed minimum level of production of somewhat over 90 percent of estimated production. The Jamaican government doubtless expected the increase in tax to be borne by foreigners. In this section we examine whether this is reasonable.

A. *Dominance*

In March 1974, seven major bauxite-producing nations (Jamaica, Australia, Guyana, Surinam, Guinea, Sierra Leone, and Yugoslavia) established the International Bauxite Association (*IBA*). Indonesia was present as an observer, and in August the Dominican Republic announced its intention to join. (See the *New York Times*, Aug. 14, 1974.) The original *IBA* members accounted for 62 percent of mine production in 1973 and control 63 percent of world bauxite reserves. Australia alone accounts for 30 percent of world reserves, followed by Guinea at 23 percent and Jamaica at nearly 7 percent. In 1973, Australia (23 percent), Jamaica (20 percent), and Guyana (10 percent) were responsible for 53 percent of world mine production (U.S. Department of the Interior, p. 15).[1] Although its members have avoided the "cartel" label and have not yet taken coordinated action, the goals of *IBA* (obtaining higher revenues and developing vertically integrated national aluminum industries) are not atypical of those of cartels, and it appears that the *IBA* could dominate the world market for bauxite if its members acted cooperatively. Meanwhile, three countries supplying much of the North American market have acted consciously parallel, if not cooperatively.

Within two months of Jamaica's unilateral actions, Guyana announced virtually identical taxes,[2] and the Dominican Republic was considering similar measures (*New York Times*, July 11, 1974). In the short term, these three nations enjoy a strong position, especially in their primary markets, even without the support of other *IBA* nations. During 1969–72, they accounted for 68 percent of *U.S.* imports of bauxite, or about 60 percent of *U.S.* mine production plus imports.[3] This fact, and others considered briefly below, suggests that these countries can export a substantial portion of their recent increases in taxes on bauxite. Most obviously, tax exporting to foreign consumers may be possible. This conclusion would be strengthened by concerted *IBA* action. Perhaps equally important, Caribbean deposits enjoy important differential resource rents which can be appropriated through fiscal measures.

Although Jamaican deposits contain only 50 percent alumina (about equal to that of Australian and *U.S.* deposits), they enjoy significant advantages of physical accessibility, silica content, and transport costs. Guyana's deposits are less accessible (owing to a heavy overburden of sand and clay) and transport costs from Guyana exceed those of most other Caribbean producers. However, the alumina content of Guyana's ore (57 percent) is the highest in the world for major deposits, and the country has a virtual monopoly on calcined bauxite (used in the abrasives and refractory industries). Discovery and exploitation of sizable new bauxite fields of comparable quality and accessibility do not appear likely to exert significant downward pressure on differential resource rents of Caribbean deposits through the medium term. The few known large, unexploited deposits are distant from the markets of Jamaica and Guyana and are located in other *IBA* nations, or in nations likely to join *IBA* (Indonesia).

The nature of alumina refining—which involves substantial weight reduction (typically 50 percent)—also generates differential locational rents, above those accruing to bauxite deposits, as such. Recent and projected developments in shipping costs are expected to increase the premia for processing alumina near extraction sites.

B. *Industrial Structure*

Six major producers dominate the international aluminum industry. They control 80 percent of western-world primary aluminum capacity, about one-half of world trade in bauxite, alumina, and aluminum, and a large share of fabrication capacity (R. A. McKern, p. B-72). The six are all fully integrated from bauxite to metal and are active in mining and processing throughout the world, often through joint ventures with one another and with smaller companies. Substantial potential thus exists for monopoly and monopsony rents from Caribbean mining activity. The existence of monopoly rents may, of course, reduce shifting to consumers, but increase tax exporting to nonresident capitalists.

---

[1] Alumina refining is less concentrated in *IBA* countries, and aluminum smelting plants are largely concentrated in developed countries near abundant, low-cost energy sources.

[2] Guyana's tax was set at about 6 percent of the ingot price. Given the higher aluminum oxide content of Guyana bauxite, the production tax per ton of bauxite virtually matched that of Jamaica.

[3] The three countries account for barely one quarter of *U.S.* alumina imports, but the share of all *IBA* members was more than 75 percent from 1969–72 (U.S. Dept. of the Interior, pp. 15–16).

C. *Capital Mobility*

In the short run quasi-rents accrue to the package of factor inputs provided by the firms engaged in the mining of bauxite—capital and the associated technological, management, and often labor skills required for modern, capital-intensive mining. Given the longevity of the substantial amounts of capital invested, (*U.S.* $800 million in Jamaica, for example; see North American Aluminum Producers in Jamaica) the economic short run may well be lengthy, and the appropriation of quasi-rents could be the vehicle of considerable tax exporting. But in the long run the multinational firms can move, especially if the tax policies of the bauxite-producing nations are not fully coordinated. Whether capital is eventually driven out will depend on the success of the *IBA* and whether taxes on bauxite go beyond merely capturing differential resource rents and monopoly rents and encroach upon the normal return to capital and know-how.

D. *Other Considerations*

Ordinary Demand and Cost Considerations

The fiscal measures considered here provide a good example of the importance of being unimportant. Bauxite costs (prior to 1974) typically have been placed at 10 percent of ingot costs (McKern, p. B-76). Therefore, the Jamaican tax increase would, if fully passed forward, raise ingot prices by only about two cents per pound, or by 6 percent, at prices prevailing in July 1974 (International Economic Policy Association, p. 82). Whether a price increase of this magnitude would induce substantial short-term shifts away from aluminum is unclear. The conductivity, strength-to-weight, corrosion resistance, and formability of aluminum suggest not. Costs of replacing aluminum in transport and electrical applications would be particularly high, but substitution would be easy in the beverage and container industry (16 percent of the market). Finally, scrap supplied about 19 percent of total *U.S.* aluminum production in 1970–73 (U.S. Department of the Interior, p. 2). Tax-induced increases in aluminum prices could serve to spur further recycling efforts. This is, however, a quite limited source of supply, except in the short run.

A major concern for the Caribbean members of *IBA* is the longer term displacement of bauxite by other abundant aluminum-bearing materials. The United States is rich in aluminum-bearing clays, such as nepheline synite, and has substantial deposits of alunite, anorthosite, and dawsonite. But the 30 to 38 percent alumina content of these ores is significantly lower than that of most Caribbean and other *IBA* deposits. If the United States were to opt to extract oil from shale, dawsonite would be a sizable byproduct, and could supply the world's aluminum needs for decades. However, potential competition from dawsonite would seem to pose little threat to *IBA* members, at least at present tax levels.[4]

Type of Tax

Because the Jamaican tax presumes a minimum level of production, it is in effect a lump sum tax if production falls below the stipulated level and is truly related to production only when the presumed minimum is surpassed. This feature prevents the majors from holding Jamaica hostage by temporarily shifting production to their mines in other countries and thereby depressing the Jamaican economy. To the contrary, Jamaica in effect is holding the firms hostage. As the tax must be paid whether bauxite is mined or not, the only economical thing to do is to mine at least the minimum amount.

---

[4] On the contrary, if *IBA* members were convinced that the United States intended to exploit its oil shale reserves, they might be well advised to raise taxes on bauxite even further, in order to extract the maximum rents before large-scale generation of dawsonite renders their bauxite deposits worthless.

E. *Summary*

Jamaica and other Caribbean members of *IBA* effectively dominate the North American bauxite market and—except for Surinam—are behaving in a consciously parallel manner. Moreover, the industrial structure is oligopolistic. Thus, in the short run in which capital is essentially immobile, we are dealing with case 6 of our conceptual framework. Over the long run, market dominance may weaken (if the tax renders Caribbean bauxite noncompetitive) or it may increase (if *IBA* becomes a strong worldwide cartel). Thus, in the long run we are dealing with either case 3 or case 9. In case 3 we would expect substantial (say, 50–100 percent) shifting to consumers, and depending upon the behavior of the firms that dominate the world and *U.S.* markets, substantial short-run shifting to consumers might occur even in case 6. Recipients of resource and monopoly rents would bear the remainder of the burden in these cases, and they and recipients of quasi-rents would bear the full burden in case 9. Given consumption patterns and the ownership of the major aluminum firms and Caribbean bauxite deposits, there is little doubt that substantial amounts of the new taxes can be exported in both the short run and the long run, subject only to constraints imposed by weakness of the *IBA* and (if *IBA* proves to be impotent) the level of differential resource rents.

Whether the *IBA* will be a viable cartel is, of course, unknowable. And because of the lack of generally recognized "arms-length" prices for bauxite, it is not possible without further detailed analysis to determine precisely whether the recent tax increases in Jamaica and Guyana have served to fully (or more than fully) wrest away the entire amount of differential resource rent accruing to their deposits. Similarly, owing to the sheer complexities of marketing and pricing and the nature of oligopolistic interaction between multinational enterprises, it is impossible to quantify monopoly rents in mining. But the behavior of one of these firms provides operational evidence that Jamaica and Guyana have not gone too far and, perhaps, that the majors expect the *IBA* to achieve its objectives. Although all the firms operating in Jamaica filed arbitration requests with the International Center for Settlement of Investment Disputes (*ICSID*) immediately after the tax increase, within six months one of the largest of the majors (Kaiser) not only agreed to abandon *ICSID* proceedings against Jamaica, but assented to the government's purchase of 51 percent of the firm's bauxite mining assets. In return, the company will receive a guarantee of reserves sufficient to maintain current rates of extraction for 30 years (*Financial Times*, Nov. 22, 1974). Though Kaiser has geographically less diversified holdings of bauxite than the other majors and is therefore under more pressure to settle, this is nonetheless a significant development.

REFERENCES

H. Hughes, "The Distribution of Gains from Foreign Direct Investment in Mineral Development," Southeast Asia Development Advisory Group, Asia Society, *SEADAG* Papers on Problems of Development in Southeast Asia, 74–10, New York 1974.

M. Manley, "Statement Before the Parliament of Jamaica on Current Bauxite Negotiations," May 15, 1974.

R. A. McKern, *Multinational Enterprise and National Resources: A Study of Foreign Direct Investment in Australian Mining Industry*, unpublished doctoral thesis, Harvard Univ. 1972.

C. McLure, "Commodity Tax Incidence in Open Economies," *Nat. Tax J.*, June 1964, *17*, 187–204.

———, "The Interstate Exporting of State and Local Taxes: Estimates for 1962," *Nat. Tax J.*, Mar. 1967, *20*, 49–75.

———, "General Equilibrium Incidence Analysis: The Harberger Model after Ten Years," *J. Pub. Econ.*, forthcoming.

International Economic Policy Association, *U.S. National Resource Requirements and Foreign Economic Policy*, Washington 1974.

Jamaica, Ministry of Finance, *Financial Statements and Revenue Estimates for 1974–75*, as presented to the House of Representatives, May 11, 1974.

North American Aluminum Producers in Jamaica, press conf., New York, May 16, 1974.

United States Dept. of the Interior, *Commodity Data Summaries*, Washington 1974.

Financial Times, London, Nov. 22, 1974.

New York Times, Aug. 14, 1974 and July 11, 1974.

U. S. Taxes on Energy Resources

By GERARD M. BRANNON*

U.S. tax policy towards energy resources is a pregnant topic. It has an interesting past, an uncomfortable present, and an unknown but promising future. Taxes, however, are not easy to separate from the whole range of government policies.

Ostensibly taxes are concerned with reducing private purchasing power, but their quality is universally measured by how they affect the allocation of resources and the distribution of income and wealth. So deeply are taxes like the whole range of government activities that in recent literature special tax reliefs have been called tax expenditures while a military draft has been described as a tax. In sum, our topic must be taxes and other energy policies.

I. Taxes Past—Pre-1974

Income Tax: The United States lightly taxes income from "producing" natural resources. The benefits from percentage depletion and the deduction of intangible drilling cost and mine development and exploration costs primarily benefit energy resources.

The tax benefits for various fuels as a percentage of the price of fuel delivered to an electric utility vary from 12 to 13 percent for oil and gas to rates of 1.4 to 4.5 percent for oil substitutes (Brannon, 1975). This ostensible discrimination would be acceptable if it could be established that there are inefficiencies in private markets which penalize production of natural resources and that can best be removed by this set of differential tax benefits. We comment on some claims for percentage depletion.

Claim (1). Private markets do not provide adequate incentive to put aside oil and gas reserve productive capacity to be used in future emergencies (because price controls and/or excess profits taxes are likely in emergencies). An extra profit incentive from lower taxes is said to provide a reserve of productive capacity for energy resources (Richard Gonzalez).

Comment (1a). There is little evidence that lower income taxes provide a reserve. Half of the tax benefits are dissipated in higher royalties (Paul Davidson, Brannon, 1974), the remainder providing a drilling incentive. The reserve issue depends on whether new drilling increases productive capacity more than production is increased. Some econometric evidence developed by CONSAD Research Corporation and by Arthur Wright and James Cox concludes that production increases *pari passu* with capacity. It has also been stated by the industry that this is the case (Mid-Continent). The only contrary evidence of which I am aware is that of Edward Erickson, S. Millsaps, and R. Spann, who find some extra reserve creation with a very long lag.

Comment (1b). More direct reserve creation methods are feasible, such as purchase of crude for storage or paying for shut-in reserves (Walter Mead and Phillip Sorenson).

Claim (2). An unintegrated corporate tax works uniquely to the disadvantage of a capital-intensive industry, which because of risk must have a high rate of return (Stephen McDonald, 1970).

Comment (2a). For large oil firms there are sufficient opportunities to spread risk (C. Jackson Grayson), so that negligible risk premium is included in the normal return. Further, a tax benefit increases supply and reduces price, which increases losses of unsuccessful firms while benefiting successful firms. This increases the variance of expected returns, an inefficient remedy for risk aversion. If risk aversion is to be regarded as a social cost to be overcome, the sensible policy would try to reduce risk, e.g., by loss guarantees.

---

* Georgetown University.

Comment (2b). If an unintegrated corporate tax discriminates against capital-intensive industries, and if this result is objectionable, the obvious solution is corporate rate reduction rather than selective relief for industries with political influence.

Comment (2c). Resource extractive industries have a lower long-run elasticity of supply than manufacturing and should, on optimal tax principles, be overtaxed rather than undertaxed (Joseph Stiglitz). The same conclusion is suggested by a general concern that our economic system does not give sufficient concern to conservation. Present tax law discourages using cheaper resources; for any end product the tax burden will be lower the greater the contribution to value added of the resources used.

Claim (3). Existing land-ownership rules make it difficult for a driller to capture the positive externalities of information which result from his drilling. This leads to some underinvestment in drilling, which can be offset by tax incentives (Edward Miller).

Comment (3a). Our land-ownership rules, specifically the absence of compulsory field unitization, result also in a negative externality and overdrilling (McDonald, 1971). There is also a negative externality in that a successful driller reduces the profit prospects of all other drillers by using up some of the limited oil and gas resources (Frederick Peterson).

Comment (3b). Information externalities can be dealt with in other ways, for example, by revising the bidding procedures on public land (Peterson).

Comment (3c). There is a bias toward overdrilling inherent in a tax law which permits a full deduction for dry hole costs. With a 10 percent success ratio, the expected cost of finding a $1 million deposit must be the incurring of $900,000 of dry hole costs. The result is that 90 percent of investment costs in exploratory oil well drilling are recovered for tax purposes within a few years. This is far more generous than the depreciation afforded other twenty-year assets.

My judgment from this argument is that the percentage depletion and the deduction of intangibles on successful wells are unjustifiable.

Another income tax energy issue is the treatment of *U.S.* companies operating abroad under high foreign tax rates. I will accept the principle that the United States should allow a foreign tax credit in the general case of foreign income of *U.S.* companies. Our law unduly favors foreign operations, however, by permitting transfer of excess tax credits on resource income to reduce *U.S.* taxes on lowly taxed foreign incomes and by too favorable treatment of startup losses. The Treasury Department proposals in 1974 dealt adequately with these problems (George Schultz).

Finally, the undertaxation of publicly owned utilities, arising from the absence of an income tax on the utility and its access to tax-free bond financing, unduly reduces electricity prices and increases resource use.

Taxes on Property: U.S. on-shore drilling is subject to severance and property taxes. To some extent severance taxes are substitutes for property taxes but they are also heavier (Allen Manvel). High severance taxes are best explained as an accommodation to the unique feature of *U.S.* property law which designates a landowner as the owner of any oil or gas that comes through a vertical well on his property. A severance tax is a crude effort to recapture a royalty interest. Since the rate does not vary with productivity, it discourages marginal operations and slightly increases prices.

Drilling rights on the outer continental shelf (*OCS*) are sold through competitive bids. This should capture the royalty, but there are questions about bidding procedures, especially the heavy front-end payments (lease bonuses) which limit competition. With the great importance of *OCS* drilling in the future, federal government royalties should recapture much of any long-term tax benefits extended to drillers.

Excise taxes: We pay for highways primarily by taxing gasoline, introducing a slight bias toward economy cars. On a comparative basis our gasoline tax is far lighter than those in other industrial countries. Total automobile and truck taxes fall short of meeting highway costs when one subtracts the revenue that would be collected by imposing regular sales taxes on gasoline. There is also an extensive literature on the inefficiency of highway financing by gasoline taxes compared to a congestion-related tax.

Semi-Taxes—Import Quotas: The United States limited oil imports from 1955 to 1973 when the world oil price was conspicuously below the *U.S.* price. This was equivalent to a tax in that it raised prices on *U.S.* oil, and it was simultaneously like a refinery subsidy since free tickets were distributed to refiners permitting them to buy stated quantities of cheap oil imports (Cabinet Task Force).

The quota system was excessively costly for the benefits derived, and it must have contributed significantly to the development of the Organization of Petroleum Exporting Countries (*OPEC*). A quota tells foreign producers that they get no more sales by reducing prices. If we had been following a reserve inventory program, we could have been in a position to expand purchases from any country willing to follow a large output, low price policy.

Semi-Taxes—Prorationing: The United States limited domestic oil production by a system of prorationing of production days per well. The system was in part conservation oriented but was also directed quite consciously at price maintenance. This, like the quotas, was simultaneously like a tax and a subsidy.

Semi-Taxes—Environmental Controls: The United States has been involved in a heavy-handed program of pollution controls. Like a tax, these controls increase energy costs. The issues involved are whether optimal tradeoffs have been achieved between benefits and costs. Considering the record of government in overestimating benefits and underestimating costs, one cannot tell if the environmental program has involved a net reduction of welfare or not.

Semi-Taxes—Natural Gas Price Regulation: Since the mid-1950's the United States has maintained control on field price on interstate sales of natural gas. Most economists who have dealt with this issue have considered that the price in terms of reduced natural gas development has been excessive as far as any benefits are concerned (MacAvoy and most of the authors assembled in Keith Brown). Regulation operates like a tax, plus a subsidy. It transfers income from producers to consumers.

II. Present Tax Issues

With current oil prices, there is no need to protect a high-cost *U.S.* industry from the rigors of competition. We have responded by trying to hold down prices. The Congress voted to roll back prices but a veto was upheld. In addition to consumer opposition to high oil product prices, there was much pressure for price control from independent refiners and marketers since high profits on crude increase the competitive power of the integrated companies.

The Administration compromised by imposing a ceiling price of $5.25 on old oil (approximately 1972 levels of production in given fields), about 50 percent above the price of a year earlier. Uncontrolled prices were allowed for new oil (increased production) plus a "bonus" amount. One barrel of old oil was freed of price control for each barrel of new production.

There has been only sparring on the tax front. The Department of the Treasury proposed only a temporary excise tax at increasing rates on the part of prices above a base of $4.50, rising to 85 percent on the part above $6.50 a barrel. The tax would phase out over five years (Schultz).

The Ways and Means Committee simultaneously adopted and gutted the Administration's windfall profits (excise) tax, while providing for the long-run phaseout of percentage depletion for oil and gas. The technique for gutting the windfall profits tax was to provide that the tax would be abated to the extent that profits were reinvested in the energy business. Limiting the incentive for energy investment to firms that had ill-gotten gains in the form of "windfall" profits serves to protect the oligopoly position of the major oil producers.

To its credit, the Ways and Means Committee bill provided for a three-year phaseout of the depletion allowance. This led to a maneuver by a group of Committee Democrats to permit a House floor vote on amendments to repeal percentage depletion for oil and gas, effective for 1974, and to eliminate the foreign tax credit on foreign oil operations. The Ways and Means Committee withdrew the oil tax bill and voted to add its provisions to a succession of more general tax reform bills which had not passed at the close of 1974.

The income distribution and resource allocation problems entailed in the massive oil price increase were left to price control, Federal Energy Administration (*FEA*) allocations, and jawbone appeals for conservation. The conservation message is pure noise, which can be roughly translated as "be efficient except where energy is involved, but if energy is involved, maximize efficiency under the assumption that the energy cost is higher than it is. As to how much higher, you guess."

These policies complicate rational energy consumption decisions. The benefit of price control is similar to a cash subsidy which rises with energy consumption. It is higher for the waster than the saver; higher for the rich than the poor. The subsidy involved in price control concentrates the increased demand on the elastic part of supply, imports. Despite official statements that the United States wants a lower world price on oil, *U.S.* policy helps keep the world price of oil up by limiting domestic producers to an average price of $8.50 for a barrel that has a marginal cost of $11.

The economic analysis behind the tax recommendations is a belief that short-run elasticity of domestic supply is low. Presumably the Treasury also believed that in the long run the elasticity of supply of oil was high, which is the implication of the recommendation to phase out the windfall profits tax over a five-year period and is a negative inference from the failure to recommend repeal of percentage depletion.

The elasticity judgment underlying the price control is that even in old wells there is an important potential for growth. A barrel of new oil can be sold at a market price of around $11. Under the Treasury windfall tax scheme, this would have borne an excise tax of about $4.50.

The effort to identify new oil is intriguing but weak (Edward Renshaw). New oil is oil produced on a property in excess of the output in the same month of 1972. This ignores the fact that, with a normal decline rate of about 10 percent a year, special efforts are required to maintain well output. The oversight is roughly offset by the rule which allows each barrel of new oil to free one barrel of old oil from controls. Assuming field output in 1974 would be 81 percent of 1972 output without new recovery efforts, the marginal revenue from oil produced over this base is $5.25 per barrel until output reaches 100 percent of 1972 and then it springs up to $11+(11−5.25) or $16.75 a barrel.

The system is made more chaotic by a Congressional action which removed price ceilings from stripper wells, i.e., wells producing under 10 barrels a day, which account for about 12 percent of *U.S.* capacity.

Continuously, the *Oil and Gas Journal* reports shortages of drilling crews, drilling equipment, oil country pipe, etc. So long as these shortages are significant, there will be little payoff from larger production incentives. For a rough test of the extent of incentive that can pay off, assume the common estimate of an elasticity of supply of 1.0 and note that drilling has increased in 1974 by 20 percent and is expected to increase in 1975 by 40 percent. If we start with a 1972 price of crude oil of $3.50 and adjust it upward for general inflation, it would not have been unreasonable to expect this sort of drilling with the incentive provided by $6 oil. An $11 price is far more than enough to produce the kind of drilling that we have seen.

This strategy shows considerable contempt for elasticity of supply of oil substitutes. The price control faces the oil substitute with an average price of oil of about $8.50, while marginal oil output is rewarded with a price up to $17. This cavalier attitude is not justified by the statement that oil substitutes have environmental problems. With high enough prices, resources can be spent to deal with environmental problems, mine safety, restoration after strip mining, etc.

A rational oil policy would start with eliminating price control. The income problems should be handled by heavier taxes on producers plus redistribution of part of the tax proceeds to consumers in a way that no one gets more for consuming more.

Politically, price control for oil and gas is attractive because it is thought mistakenly that this is the best way to avoid the massive income redistribution from consumers to producers that is involved in large price increases. In economic policy, however, income objectives are far better attained by using income tools than by using price tools.

The first step to increase taxes on producers of oil and gas is to repeal the income tax benefits, percentage depletion, and the current deduction of intangible drilling expenses on successful wells.

It follows from our previous argument about short-run elasticity that an $11 price of crude involves short-term windfalls, i.e., income in excess of the amount necessary to call forth the feasible increase in output over the next few years. As a technical matter, an excise tax on windfall profits could be constructed to parallel exactly the price control of $5.25 a barrel on old oil. A producer should be indifferent between a law that says he must sell an $11 barrel of oil for $5.25 and one that permits sale at $11 with a tax of $5.75.

A little thought could produce a windfall profits tax that would be more efficient than the present price control. Design of an optimum windfall tax would depend primarily on a detailed description of the elasticity of supply. Broadly, the distinction between the new oil and old oil seems useful, although it is our judgment that there is no justification for as large a net difference in the allowable net prices as that provided under present price control rules.

The windfall tax needs to be gradually phased out. In the long run there is, however, an analogous windfall income which would be reflected in royalties. With respect to *OCS* production, this can be recaptured by government bidding procedures for drilling rights. With respect to on-shore production, this is partly recaptured by severance taxes. The latter recapture could be more effective if severance taxes were moderated toward being more precisely adjusted to the royalty on each property, reaching zero on the marginal property—shades of Henry George.

Some very rounded numbers will permit a general quantitative estimate of the revenues under discussion. Domestic crude production is about 3.5 billion barrels a year, 60 percent of which is old oil. I assume the windfall profits tax replicates the producer burden of present price control, although it could be higher. I estimate the values of the present income tax benefits on $5.25 oil are a deduction of 18 percent for percentage depletion and 8 percent for intangibles, and on $11 oil, 20 percent for percentage depletion and 6 percent for intangibles.

The tax benefit in 22 percent depletion is cut down in value by various legal offsets which become less as oil prices rise (Brannon, 1975). Also, at a price of oil of $3, the long-run value of the deduction of intangibles is about half that of percentage depletion (Susan Agria, Brannon, 1975).

The assumptions provide the following revenue gains:

(1) the increased revenue from removing percentage depletion and the intangible deduction on successful wells on 3.5 billion barrels of oil at $11, i.e., (.48)(.26)($11) (3.5 billion) amounts to $4.8 billion;

(2) the windfall tax on 2.1 billion barrels of old oil at $5.03, i.e., $5.75 less differential income tax benefits (2.1 billion)(5.75)(1−(.48)(.26)), comes to $10.6 billion; and

(3) the total is $15.4 billion.

Increased cost to the public for oil would be $5.75×2.1 billion barrels or $12.1 billion. If this much of the revenue were refunded to consumers in the form of a refundable tax credit of $60 a person ($240 for a family of four), consumers on the average would have been made whole. (State experience with refunds of sales and property taxes of the poor show that a program like this is feasible.)

In the long run this refund would have to be moderated to the extent that we had let prices rise above the present price control limits.

The same general approach is applicable to natural gas.

The percentage depletion deduction and the deductibility of certain exploration and development expenses should be terminated for all resources, including the other energy resources coal, uranium, and geothermal steam.

III. Future Taxes

With the present high world price, Project Independence has captured the imagination of the public and the politicians alike. There are a number of plausible concepts of energy independence, such as (1) a zero level of imports and (2) reserves adequate to cover imports reduced by a feasible level of reduced consumption.

Before I comment on the choice of policies, note that concept (1) could be defined equally well by the expression "*OPEC* pricing itself out of the American market." If this occurs, the independence problem solves itself. At present world oil prices, this seems plausible. If we adjust the pre-embargo *U.S.* oil price of $3.50 generously for inflation to put it at $4.50, the current import price is about a 150 percent increase. As of 1972, standard long-term projections were that the United States would provide about half of its oil supply in the 1980's (National Petroleum Council). It is also estimated that the long-run supply elasticity is about 1.0 (Henry Steele, Schultz). Assuming this historical elasticity prevails in the future, *U.S.* production should double and fully replace imports. Extrapolation is hazardous, but there is good reason to expect that at prices of $8 to $10 new supplies from shale, or coal liquefaction, or expanded nuclear energy could increase supply elasticity at high prices. At an $11 price, there should be significantly less oil demanded. Long-run demand elasticity has been estimated at 0.5 (J. C. Burrows and T. A. Dommenich).

While independence is plausible even without special government policies, the prospect left out of account is the *OPEC* cartel strategy. A substantial development of alternate energy sources in the United States, Europe, and Japan should push the cartel to a lower price strategy.

If we develop alternate *U.S.* sources to provide a substitute for high foreign prices, the foreign prices are likely to come down, leaving the expanded *U.S.* industry in a depressed state that cries for protection. If the prospects of future losses from lower foreign prices deter *U.S.* development, then lower foreign prices will not materialize.

The present *U.S.* policy position is an odd combination of heavy subsidies for research and development on alternate sources, principally nuclear, plus a price control position on oil that, as we noted earlier, provides generous marginal prices for oil development but abnormally low average oil prices to deter current development of all nonoil substitutes.

At this point, a most effective form of intervention would seem to be one that goes to the problem of prospective losses from future reductions in world oil prices. A constraint on this policy should be, however, that the problem be solved in a way that will not deprive *U.S.* consumers of any advantages that might accrue from lower world oil prices.

This specification does seem to eliminate the kind of oil import quota that we maintained in the period of the 1950's to 1973. A sensible kind of policy would seem to be one that was willing to accept oil imports at any time. If unreliable imports create security problems, then a contingency reserve should be maintained, financed by a tax on imports. The added security is a service provided to users of imports and represents a real cost that importers should pay. The size and character of a contingency reserve would have to be determined from a probability evaluation of various interruption scenarios. It is commonly asserted that a reserve equivalent to one year's imports is adequate, and this would involve a continuing cost of $1 per barrel. (Mead and Sorenson). The cost is annual interest and maintenance, so a continuing import tax is needed.

A strategy for building up this reserve could be an open-ended offer to buy in bulk from any producer country that would make a significantly lower price available.

Beyond this contingency reserve financed by import taxes, we have to encourage development of alternate sources to "break" the cartel. The *U.S.* government should deal with this by minimizing prospective losses to producers of substitutes, but this should be done economically. This suggests that the *U.S.* commitment be limited to a loss guarantee provision similar to provisions built into war procurement contracts to cover a sudden end of the war. The loss guarantee should be designed to cover the part of fixed costs that could not be covered by continued operation in the face of future imported oil prices (that would include the security tax of about $1). Unlike a subsidy in the form of a return to oil quotas, the loss guarantee policy would be designed to cover only sunk costs and not to encourage continued operation of high-cost domestic substitutes that would involve mounting losses.

Project Independence must be designed so as to avoid its becoming an old-fashioned protective tariff.

REFERENCES

S. Agria, "Special Tax Treatment of Mineral Industries," in A. Harberger and M. Bailey, eds., *The Taxation of Income from Capital,* Washington 1969.

G. Brannon, *Energy Taxes and Subsidies,* Cambridge, Mass., 1974.

———, "Existing Tax Differential and Subsidies Relating to the Energy Industries," in G. Brannon, ed., *Studies in Energy Tax Policy,* Cambridge, Mass., forthcoming 1975.

K. Brown, ed., *Regulation of the Natural Gas Producing Industry,* Resources for the Future, Baltimore 1972.

J. C. Burrows and T. A. Dommenich, *An Analysis of the U.S. Oil Imports Quota,* Lexington 1970.

J. Cox and A. Wright, "The Cost Effectiveness of Federal Tax Subsidies for Petroleum Reserves: Some Empirical Results," in G. Brannon, ed., *Studies in Energy Tax Policy,* Cambridge, Mass., 1975.

P. Davidson, L. Falk, and H. Lee, "The Relations of Economic Rents and Price Incentives to Oil and Gas Supplies," in G. Brannon, ed., *Studies in Energy Tax Policy,* Cambridge, Mass., 1975.

E. Erickson, S. Millsap, and R. Spann, "Energy Policy and the Depletion Allowance," pap. prepared for the Brookings Panel in Econ. Activity, 1974.

R. Gonzalez, "Percentage Depletion for Petroleum Production," Hearings on Tax Revision, Ways and Means Comm., U.S. Cong., 1957.

C. J. Grayson, Jr., *Decisions Under Uncertainty: Drilling Decisions by Oil and Gas Operators,* Boston 1969.

A. Manvel, "A Survey of the Extent of Unneutrality Toward Energy Under State Excise, Property and Severance Taxation," in G. Brannon, ed., *Studies in Energy Tax Policy,* Cambridge, Mass., 1975.

S. McDonald, *Petroleum Conservation in the U.S.,* Baltimore 1971.

———, "Distinctive Tax Treatment of Income from Oil and Gas Production," *Nat. Res. J.,* Jan. 1970, *10,* 97–112.

W. Mead and P. Sorenson, "A National Defense Petroleum Reserve Alternative to Oil Import Quotas," *Land Econ.,* Aug. 1971, *47,* 211.

E. Miller, "Some Implications of Land Ownership Patterns for Petroleum Policy," *Land Econ.,* Nov. 1973, *46,* 413–23.

F. Peterson, "Two Externalities in Oil Explorations," in G. Brannon, ed., *Studies in Energy Tax Policy,* Cambridge, Mass., 1975.

E. Renshaw, "The Pricing of Domestic Crude Oil," paper presented at meeting of East. Econ. Assn., Albany, N.Y., 1974.

G. Schultz, Secretary of Treasury testimony, Oil Profits Hearing of Comm. on Ways and Means, Washington 1974.

H. Steele, "Cost Trends and the Supply of Crude Oil in the U.S.," in M. Macrakis, ed., *Energy,* Cambridge, Mass., 1974.

J. Stiglitz, "The Efficiency of Market Prices in Long-Run Allocations in the Oil Industry," in G. Brannon, ed., *Studies in Energy Tax Policy,* Cambridge, Mass., 1975.

Cabinet Task Force, *The Oil Import Question,* Washington 1969.

CONSAD Research Corp., "The Economic Factors Affecting the Level of Domestic Petroleum Reserves," in *U.S. Treasury Dept., Tax Reform Studies and Proposals,* Washington 1969.

Mid-Continent Oil and Gas Assn., "Analysis and Comment Relating to the CONSAD Report on the Influence of U.S. Petroleum Taxation on the Level of Reserves," Tax Reform, 1969, Hearings before the Senate Finance Comm., 91st Congress, 1st Session, Part 5, 4627–47.

National Petroleum Council, *U.S. Energy Outlook,* Washington 1972.

DISCUSSION

ARTHUR W. WRIGHT, University of Massachusetts: Both the Gerard Brannon and the Malcolm Gillis-Charles McLure papers illustrate the point that we cannot deal adequately with the taxation of natural resources in isolation from alternative, nontax policy measures. By recognizing this point, Brannon is able to put energy taxes in proper perspective and to suggest interesting interpretations of nontax policies (e.g., the oil import quota). In contrast, by focusing on tax incidence in isolation from other, more basic policy issues, Gillis and McLure do not fully exploit the potential of their topic.

It is appropriate to begin the discussion of Brannon's paper by indicating substantial agreement with his treatments of (1) the percentage depletion allowance and the expensing of "intangible" drilling and dry-hole costs; (2) the price controls on domestic crude oil and natural gas; and (3) the proposed loss-guarantees for high-cost alternatives to *OPEC* oil. There is less agreement on foreign tax credits.

To me, the most interesting part of Brannon's paper is his treatment of a windfall profits tax on oil and natural gas. As a preliminary, we can improve upon his definition of "windfall income." According to that definition, firms would realize windfall profits only when they were operating along the perfectly inelastic portions of their supply schedules. In fact, windfall profits can be said to arise from any unexpected increase in market price. The quarrel here is more than semantic: under Brannon's definition, nothing varies but the distribution of the windfall rents, whereas under the broader definition attempts by governments to appropriate those rents may affect market price and quantity—that is, the efficiency of resource allocation. Unless we want to discuss Alfred Marshall's "market period" case (and no bureaucracy, let alone a legislature, could ever move fast enough to appropriate market-period rents), in the interesting cases both demand and supply will have finite, nonzero slopes.

The windfall profits excise tax measure proposed by the Treasury, which Brannon apparently favors, would raise price and reduce quantity compared to the no-tax equilibrium. In the familiar story, the resulting tax revenues and deadweight loss would be borne by both consumers and producers. When the Treasury first mooted its proposal, the *OPEC* cartel and the Arab embargo had already inflicted upon consumers some deadweight loss and transfers of wealth to foreign and domestic producers; the Treasury's excise tax would compound both the inefficiency and the redistribution of income away from consumers. The tax revenues could, of course, be rebated to consumers, as Brannon recommends, but the added inefficiency would remain.

By the rough numerical estimates presented at the end of section II of Brannon's paper, consumers could be more than compensated out of tax revenues for the increase in price resulting from the excise tax. This result is, however, specific to the particular assumptions made in the calculation (which may be suggestive but is too rough to serve as the basis of a major policy initiative).

Brannon advocates gradually phasing out a windfall profits tax according to a definite timetable; however, he dismisses too cavalierly the possibility that such a scheme would induce destabilizing intertemporal shifts in supply. It would be worth exploring whether uncertain timing of tax reductions might not prevent or discourage such shifts. (This is not an argument for the Brownian motion of statements and retractions which passes for energy "policy" in the Ford Administration.) Interestingly, it would also be worth exploring whether an announced timetable for phasing out percentage depletion and the special expensing provisions for oil and gas corporate income might not give an added incentive to produce more petroleum now, thereby easing the present burden on consumers *and* cutting windfall profits in the industry. In the topsy-turvy world of the energy crisis, no paradox is implausible.

The foregoing analytical complexities, plus the apparent impracticability of differentiated taxes on economic rents, suggest that it is not self-evident that "a little thought could produce a windfall profits tax that would be more efficient than the present price control." This may explain why governments so readily resort to price controls. Not that controls are the right answer, as Brannon persuasively argues. But why not then (as he also argues) rely on income tools rather than price tools to attain income objectives?

The possibility of an *OPEC*-like cartel in the world market for bauxite is a highly interesting subject. Unfortunately, Gillis and McLure focus on a relatively uninteresting aspect of that subject: the incidence of any taxes which Jamaica and other bauxite-producing countries may impose on the aluminum companies is distinctly not the focal issue in the world bauxite market. Thus their framework for analyzing tax incidence, which occupies the first half of the paper and which I find confusing and cumbersome rather than helpful, is not really used in the second half of the paper on Jamaica's new bauxite tax. This latter section of the paper is tantalizing but, in the end, not satisfying.

The central issue can be posed quite simply: *without* a world bauxite cartel, Jamaica and other members of the International Bauxite Association (*IBA*) can appropriate at most the total rents accruing to producers in a competitive market; *with* a cartel, however, they can appropriate all that and much more—namely, monopoly rents at the expense of consumers and world economic efficiency. Assume for simplicity that foreigners now own all the rights to Jamaican bauxite, and that Jamaicans now consume a negligible proportion of world output of aluminum products. Then, without a cartel, Jamaica can tax away (or otherwise secure) producer rents at the expense of the current foreign owners. With a cartel, however, Jamaica will be able to reap in addition a transfer of wealth from foreign consumers of aluminum products and also to foist upon the non-bauxite-producing world some amount of deadweight loss. So much for the "incidence" of different pure cases in the bauxite market; the interested reader can lay out the mixed cases on his own.

The crucial question here is whether the *IBA* will follow successfully in *OPEC*'s footsteps. Gillis and McLure give us several clues in the second part of their paper, although we still lack sufficient information to answer that question with confidence. The structure of the new tax—the stipulated minimum output and the tie to ingot prices—is reminiscent of the "posted prices" used to compute the tax liabilities of the oil companies in the Middle East. The early capitulation of Kaiser Industries (and, most recently, of Alcoa in Surinam) is reminiscent of the cooperative attitude of the oil companies towards *OPEC*'s price increase goals. Yet a nagging doubt remains that there are too many competing sources of alumina, over too wide an area and under too numerous a set of jurisdictions, for the *IBA* to achieve and maintain the internal discipline displayed to date by the largest oil exporters in *OPEC*.

Gillis and McLure hint at a number of other possible paper topics, some of them involving tax issues, dealing with the world bauxite market. For example, assuming no cartel, what set of taxes or other policy measures (e.g., nationalization) should Jamaica adopt in order to appropriate the maximum producer rents? What set of individual and joint measures would maximize the *IBA*'s chances of forming a cartel? What does the possibility of a bauxite cartel imply for *U.S.* policy towards oil shale, which contains the low-grade source of alumina, dawsonite, and which is also affected by the *OPEC* cartel? Should, or can, the United States alter its foreign tax credit laws to prevent the *IBA* from siphoning off rents which could go to the U.S. Treasury? One's appetite is whetted, but the plate is, as yet, still empty.

OPTIMIZING MODELS OF PUBLIC DECISION MAKING

Purposive Models of Legislative Behavior

By JOHN A. FEREJOHN AND MORRIS P. FIORINA*

Although political scientists have studied Congress and other legislative bodies for many years, they have been relatively unsuccessful in developing a general theory of legislative behavior. Picking and choosing from the conceptual stores of sociology and psychology, political scientists have organized their research around concepts such as role, norm, integration, etc. Almost without exception the resulting partial theories are quite vaguely formulated, so that one cannot determine which propositions strictly derive from the theory. In other words, though we have good descriptive information about how certain legislatures work, we have a very limited set of theoretical propositions that can help to explain these workings. Still, some scholars appear to have excellent intuitions about how alterations in various internal or external institutions affect the operation of legislatures and their policy output. Thus, we believe there is a deeper understanding of legislative behavior than the lack of a theoretical superstructure suggests.

In recent years several developments have shown some promise of filling the theoretical gaps in the study of legislative behavior. First, game theorists and social choice theorists have begun to formulate abstract models of simple legislatures. Thus we now have a short but growing list of propositions about legislative behavior and policy making in certain highly simplified situations. Second, some empirical researchers have gradually become aware that legislators are goal-seeking agents who choose from available strategic alternatives to further their ends. Some recent empirical work reflects this realization. Third, a number of scholars have been working to bridge the gap between highly abstract social-choice and game-theoretic models of legislatures on the one hand and detailed empirical studies of legislatures on the other. In this necessarily short review we will provide a brief survey of the work underlying each of these developments. The discussion which follows proceeds in stages from the abstract constructions of the game theorists to the more concrete descriptions of empirical researchers.

I. Game Theory and the Legislative Process

Since the publication of *The Theory of Games and Economic Behavior*, several authors have proposed simple game-theoretic models of legislatures. In this section we describe the basic elements of such models and discuss the kinds of propositions they imply. Game-theoretic models typically have a set of *actors*, each of whom have *preferences* defined over a set of "alternatives" or "social states." These preferences are usually assumed to satisfy certain rationality properties. Subsets of the set of actors are called *coalitions* and coalitional preference is usually defined as follows: coalition is said to prefer x to y if and only if each of its members prefers x to y.

A principal notion in game-theoretic

---

* California Institute of Technology.

analysis of legislatures is the idea of coalitional power. One way to formulate this notion is as a characteristic function. The characteristic function is a mapping from coalitions into subsets of the alternatives, which specifies for each coalition in its domain the alternatives that coalition can enforce (S. D. Bloomfield and R. B. Wilson). Another way of conceptualizing coalitional power is a set of binary *effectiveness* relations, one for each coalition, which describe for each coalition its ability to determine the outcome of the legislative process on pairs. That is, if a given coalition is effective for x against y, then it can ensure that if the legislature is choosing between x and y, x will be chosen. A game is simply the set of players and the set of alternatives together with a list of permissible or formable coalitions and some description of the powers of the formable coalitions.

The three elements introduced thus far (*preferences* held by the members, formable *coalitions*, and *powers* of each of the formable coalitions) enable us to make some statements about what a legislature will do when faced with particular choices. One can define a dominance relation as follows: x dominates y if and only if there is a permissible coalition that can ensure x and which prefers x to y. In the next section we will review the work of a number of scholars who have placed sufficient restrictions on the distribution of legislator preferences to ensure the existence of an undominated element. As is well known, however, without such restrictions many legislative institutions cannot be counted on to produce a dominance relation which has an undominated element. This has led some scholars to theorize about what will happen if no undominated element exists.

Several authors have developed solution concepts that allow one to make assertions about what will happen in games without undominated elements (G. Owen). Unfortunately, the solution concepts advanced thus far typically do not make very strong statements about the outcomes of a game. In many games of interest the solutions constitute a very large set relative to the set of possible outcomes.

As a result William Riker suggests that while we may be unable to predict what alternatives will be chosen in a given legislative situation, we may be able to make assertions about which coalitions will form and/or remain stable. He proposes a rule—the size principle—which asserts that in constant-sum games of perfect information only minimal-winning coalitions form. This principle has received careful attention from theorists (Robert Butterworth, Kenneth A. Shepsle 1974a, 1974b, Richard McKelvey), but its status as a theoretical proposition remains ambiguous. For, except in noncontroversial cases in which an undominated element exists, each alternative is dominated by some other, so that even if a minimal-winning coalition moves the legislative from y to x, another coalition could move the legislative from x to z. In general, one cannot expect coalitions to be stable. Perhaps Riker is using the notion of coalition formation in a somewhat different sense than is natural in a game-theoretic terminology. One might define a somewhat different concept of coalition formation and argue that even when the domination relation is cyclic only minimal-winning coalitions form. Recently McKelvey has shown that, if a certain restriction is placed on the payoff configurations that are admissible, a version of the size principle follows. As a descriptive principle, however, the condition McKelvey proposes is dubious.

A number of political researchers have conducted empirical studies to see how accurate the size principle is as a descriptive hypothesis. Some examples are those by David Koehler, Barbara Hinckley, and David Moore. None of these authors finds a pronounced tendency for minimal-winning coalitions to form in legislative voting behavior. Thus, the size principle does not appear to describe legislative coalition formation, at least as measured in these studies. Nonetheless, despite its limited theoretical and empirical support, many scholars (of both theoretical and empirical inclinations) find the size principle intuitively appealing.

II. Spatial Models of Legislatures

In spatial models of legislatures the set of alternatives is assumed to be a subset of Euclidean space, and each legislator is assumed to have a most preferred point within the subset. Legislator preferences may be positively monotonic in distance (Duncan Black), convex, and representable by a differentiable utility function (Charles Plott), convex (Judith Sloss), and/or separable (Gerald Kramer). The basic idea of this type of models is to provide conditions on preferences sufficient to ensure the existence of an undominated point in the set of alternatives.

Black proves that if preferences are convex (i.e., single peaked) in a one-dimensional Euclidean space and if each person votes for the alternative he prefers in each pairwise contest, then there is an alternative (the median position) that cannot be defeated in a pairwise majority vote. As is well known, if the dimensionality of the space exceeds one, convexity is no longer a sufficient condition for the existence of an undominated element.

Plott provides a set of sufficient conditions for the existence of an undominated element when the set of alternatives is a subset of an n-dimensional Euclidean space, and each voter has a concave, differentiable utility function on the set. Plott's condition roughly may be stated as follows. If there is an odd number of voters and they are arranged in a subset of Euclidean space in such a way that one voter is located at an interior point, x, then x is an undominated point if and only if for each utility gradient pointing in a given direction there is another pointing in the opposite direction. Sloss generalized Plott's theorem to weaken the requirement that voters have differentiable utility functions. She requires only that voter preferences be quasi-transitive and satisfy a certain convexity property. In her work convex cones with "gradientlike" properties play much the same role as the gradients in Plott's theorem.

Kramer extends results of Black, A. K. Sen, and others to the case of a multidimensional space of alternatives by giving a class of sufficient conditions for the existence of an undominated element. He argues that the class of conditions which he and Plott provide are so restrictive they are "unlikely to be satisfied in practice except by care and lucky accident" (Kramer 1973, p. 296).

In another contribution Kramer (1972) shows that if legislators have convex and separable preferences over a subset of an n-dimensional Euclidean space, if motions to amend the status quo, x, can change the status quo in at most one component (i.e., issue) at a time, and if legislators vote for a motion if and only if they are made better off by its passage, then there exists a point which is undominated. That is, if legislators do not cooperate too extensively in making their voting decisions, the rules restricting the agenda may create undominated elements. Of course such procedural devices will be ineffective if legislators coordinate their voting strategies sufficiently. Nevertheless, such coordination may be costly enough that the conditions of the theorem are sometimes met in practice.

III. Vote Trading

Some authors have suggested that vote trading is a desirable activity in legislatures and that it may lead to outcomes socially superior to sincere voting outcomes (James Buchanan and Gordon Tullock). Numerous papers on vote trading have been published, but one finds a notable lack of consensus among their authors both on definitions and conclusions. The most important accepted result is the following: if any trade may occur among legislators, and if legislator preferences are separable, a point, x, is an equilibrium in the vote-trading process if and only if it is the outcome of issue-by-issue voting with no trades occurring. Of course the trading process may have no equilibrium. (See Koehler 1975, Peter Bernholz, Joseph Kadane, and Ferejohn 1974b.) The basic idea of the proofs is simply that if any coalition may be organized costlessly to carry out a trade, and if the ultimate decision rule is the method of majority decision, then informal or formal institutions may always be circumvented by an appropriately organized majority. Only to the extent that trades are costly to implement and to police can formal or informal institutions create undominated elements where none would exist under simple majority rule with sincere voting.

We think that this observation may be of considerable importance in understanding how an institution like the Congress works. The Congress is not merely a bicameral body with decisions in each house made by simple majority rule. Rather, there are a number of formal and informal institutions (the committees, the parties, the state party delegations, the leadership systems, the policy groups, etc.) which hold more or less regular places in the legislative process. The presence of such institutions makes some coalitions cheap to organize and others much less so. Differential organization costs may affect not only the existence of equilibria, but also the distributive consequences of policy outcomes (Ferejohn, 1974a). While little theoretical or empirical work has taken place along these lines, it appears that members of certain institutions within the Congress have distinct advantages with respect to particular legislative outputs. The existence of regular distributional effects is difficult to explain without some notion that certain coalitions are more difficult to form than others (namely those that could upset the allocation in question).

IV. Models of the Committee Assignment Process

While legislation may be proposed and amended on the floor of either chamber of the Congress, in practice this almost never happens. Instead, a proposed bill is referred to a committee, which then decides whether the bill will be reported to the whole chamber and, if so, in what form. As a rule, few amendments are offered on the floor (although committee to committee variation is wide) and even fewer are successful. But although most of the policy output of the Congress is determined in committee, little attention has been devoted to incorporating the standing committee system into models of the Congressional process. Recently, however, theorists have focused on one aspect of the committee system—the assignment process. Using simple optimizing models they can successfully account for most of the available data about Congressional committee assignments. Unlike models which require only that legislators be rational agents (i.e., that they be maximizing *something*) the assignment models postulate a specific objective function which the Committee on Committees tries to maximize.

The assignment process operates as follows. Newly elected Congressmen and those dissatisfied with their current assignments submit requests to their party's Committee on Committees. These requests take the form of partial rankings of the committees of the House. The Committee on Committees then meets to consider these requests and make the actual assignments. There are some formal restraints on assignments: all members must receive a committee; members may be placed on one exclusive committee, or on one semiexclusive and one nonexclusive committee, or on two nonexclusive committees. Additionally, the descriptive literature identifies various informal constraints on assignments: no freshmen on exclusive committees; reserved seats for certain states on certain committees; higher priority for nonfreshmen requesters than for freshmen.

In Shepsle's (1975) model of the assignment process, both the informal constraints and the ranking information about requests are omitted. Instead, Shepsle assumes that the Committee on Committees maximizes the number of requests satisfied. Using data only for freshmen requesters and for the Democratic Committee on Committees for four Congresses, Shepsle finds that this extremely simple model accounts for "most" of the assignments in the following sense: while it might be possible to satisfy a few more requests by rearranging assignments, the number of feasible assignments is so large and the search costs sufficiently high that the actual assignments seem a reasonable approximation to the optimum.

Michael Cohen's model employs a similar objective function. Interestingly, the Committee on Committees allegedly considers the committees in alphabetic order. After studying the consequences of various sequential assignment methods, Cohen concludes that quite different assignments would result if assignments were considered in other sequences.

The relatively good fit of the Shepsle and Cohen models suggests that committees are made up largely of self-selected members. Farm state Representatives migrate to Agriculture; Western Representatives gravitate to Interior. Given that the committees exercise a great deal of influence in writing legislation, the end result is a situation in which policy advocates write public policy for the whole Congress. In an interesting work William Niskanen examines the consequences of a system in which committees dominated by program advocates write legislation; he concludes that such a system has undesirable allocative properties. Ferejohn's (1974a) empirical study of Congressional policy making in the area of water resources supports Niskanen's conclusions.

V. Models of Roll Call Voting

One of the most conspicuous aspects of legislative behavior is the roll call vote. The voting records constitute a happy hunting ground for statistically oriented political scientists, and a few models of the vote decision now exist.

Duncan MacRae first proposed a model of roll call voting. The model is theoretically quite ambitious: it attempts to integrate a spatial model of electoral behavior in the constituency with a spatial model of roll call voting in the legislature. Representatives do not simply take the median position of their districts in this model because their votes depend on their personal policy preferences, their electoral margins, and the overall "political complexion" of their districts, in addition to their constituents' approval and disapproval. Given the relative complexity of the model, general results are somewhat hard to come by.

In another contribution, Fiorina (1974) views Representatives as using their vote either to maximize their probability of reelection or to maintain it at some personally satisfactory level. In this model Representatives make decisions under uncertainty about the likelihood that their vote will elicit an electoral reward or punishment of a given magnitude. The homogeneity of a district's group structure emerges as critical in Fiorina's model. Optimal roll call voting strategies always exist in homogeneous districts, but not in heterogeneous districts. Other propositions in the model concern the effects of (1) variations in rewards and punishments stemming from votes, (2) the likelihood a vote will become a campaign issue, and (3) abstention.

In contrast, Jane Gilbert offers a decision-making-under-certainty model of Congressional roll call voting. She assumes Congressmen desire only 50 percent $+\epsilon$ of the vote. In the model, this limited electoral objective frees them to cast other votes with party leaders in order to build up credit within the institution. The number of free votes Congressmen have depends on two dimensions of district heterogeneity: consensus and diversity (the former refers to agreement on a single roll call, the latter to agreement across a series of roll calls). Interestingly, Gilbert's conclusions about the effects of district heterogeneity directly contradict Fiorina's. In the Gilbert model, Congressmen wish to have diverse districts—they should be willing to ask state legislatures to redistrict them out of safe, homogeneous districts. This startling difference between the two models stems largely from the differing candidate-objective functions assumed. In Gilbert's model candidates strive to get only one vote more than a tie. If ϵ were set higher, say 10 percent, the conclusions of the Gilbert and Fiorina models would be similar.

VI. Mainstream Political Science and Optimizing Models of Congress

Of the models thus far discussed, only a minority have been the creation of political scientists. As a discipline, political science has been influenced far more by sociology and social psychology than by economics. Research on Congress, in particular, is organized around concepts such as "role," "norm," and "socialization" rather than "preference," "choice," and "maximization." But, very recently, mainstream Congressional scholars have given strong indications that they are broadening their collection of intellectual tools.

For example, Richard Fenno presents a large-scale comparative study of Congressional committees within the framework of a purposive analytical model. Fenno sees Congressmen as desiring reelection, good public policy, and institutional influence—different mixes for different Representatives. The committees, in turn, offer differing opportunities to satisfy these differing goals (e.g., Public Works or Interior are clearly more valuable for electoral purposes than Foreign Affairs). Fenno argues that Congressmen's goals affect the committee assignments they seek and, in conjunction with the political environment, the formal structure and informal procedures of the committees.

Other empirical works which show the utility of purposive models of behavior include James T. Murphy's analysis of the House Public Works Committee decision making, John F. Manley's study of the House Ways and Means Committee, Barbara Deckard's analysis of state delegations in the House of Representatives, David Rohde's and Shepsle's examination of committee requests in the House of Representatives, John Kingdon's excellent study of roll call voting decisions, and David Mayhew's interpretative review of the Congressional literature.

Empirical research on legislatures has yielded more than detailed descriptions of Congressional institutions, processes, and behavior. It also has provided a growing list of regularities which seem sturdy enough to bear close theoretical examination. Some of these observations are as old as the hills, but receive support from recent work as well: the work of the Congress is done in committees; committee decisions are seldom overturned on the floor; party membership is the single most important predictor of roll call votes; certain committees (such as Appropriations and Ways and Means) are consistently more successful on the floor than are others (Education and Labor or Agriculture, for example). There are other such propositions which could be culled from an examination of the empirical literature. To a surprising extent these propositions appear to hold not merely in the House and Senate, but also in many state legislatures as well.

The existence of empirical regularities in a relatively wide variety of legislative situations (across time as well as cross-sectionally) is a challenge to theorists. What is it about legislative bodies that makes committee decisions tend to remain stable during floor consideration under simple majority rule? The empirical literature on Congress indicates not only that seemingly undominated alternatives exist, but that these alternatives regularly emerge as committee bills. The theoretical literature suggests that this is an extraordinarily improbable state of affairs. Surely we should search for an explanation of this phenomenon.

In our opinion, the theoretical literature on legislative bodies is sufficiently advanced that we may begin to hope for answers to the preceding kind of question. One would hope that over the next few years theorists will take the body of empirical research seriously enough to attempt to provide explanations for observed regularities, and simultaneously to provide a theoretical base for future empirical research.

REFERENCES

K. J. Arrow, *Social Choice and Individual Values*, New York 1951.

P. Bernholz, "Logrolling, Arrow Paradox and Cyclical Majorities," *Pub. Choice*, 1973, *15*, 87–96.

D. Black, *The Theory of Committees and Election*, Cambridge 1958.

S. D. Bloomfield and R. B. Wilson, "The Postulates of Game Theory," *J. of Math. Soc.*, 1972, *2*, 221–34.

J. Buchanan and G. Tullock, *The Calculus of Consent*, Ann Arbor 1962.

R. Butterworth, "A Research Note on the Size of Winning Coalitions," *Amer. Polit. Sci. Rev.*, 1971, *65*, 741–45.

M. D. Cohen, "The Importance of Member Preferences in Committee Assignment: An Assessment Against Optimal Standards and a Simple Process Model," presented at MSSB Conf. on Math. Models of Cong., Aspen 1974.

B. Deckard, "State Party Delegations in the U.S. House of Representatives—A Comparative Study of Group Cohesion," *J. of Polit.*, 1972, *34*, 199–222.

R. F. Fenno, Jr., *Congressmen in Committees*, Boston 1973.

J. A. Ferejohn, *Pork Barrel Politics: Rivers and Harbors Legislation, 1947–1968*, Stanford 1974a.

———, "Sour Notes on the Theory of Vote Trading," Calif. Instit. of Tech. Soc. Sci. work. pap. no. 41, 1974b.

M. P. Fiorina, *Representatives, Roll Calls, and Constituencies*, Lexington 1974.

J. Gilbert, unpublished doctoral dissertation, Univ. of Rochester 1975.

B. Hinckley, "Coalitions in Congress: Size and Ideological Distance," *Midwest J. of Polit. Sci.*, May 1972, *16*, 197–207.

J. B. Kadane, "On Division of the Question," *Pub. Choice*, 1972, *13*, 47–54.

J. W. Kingdon, *Congressmen's Voting Decisions*, New York 1973.

D. H. Koehler, "The Legislative Process and Minimal Winning Coalition," in Richard G. Niemi and Herbert F. Weisberg, eds., *Probability Models of Collective Decision Making*, Columbus 1972, 149–64.

———, "Vote Trading and the Voting Paradox: A Proof of Logical Equivalence," *Amer. Polit. Sci. Rev.*, forthcoming 1975.

G. H. Kramer, "Sophisticated Voting Over Multidimensional Choice Spaces," *J. of Math. Soc.*, July 1972, *2*, 125–80.

———, "On a Class of Equilibrium Conditions for Majority Rule," *Econometrica*, 1973, *41*, 285–98.

D. MacRae, Jr., *Dimensions of Congressional Voting: A Statistical Study of the Eighty-First Congress,* Berkeley 1958.

J. F. Manley, *The Politics of Finance: The House Committee on Ways and Means,* Boston 1970.

D. R. Mayhew, *Congress: The Electoral Connection,* New Haven 1974.

R. McKelvey and R. Smith, "A Comment on the Debate Over Riker's 'Size Principle'," presented at the MSSB Conf. on Uncertainty, Polit. Processes and Pub. Policy, San Diego, Aug. 1974.

D. W. Moore, "Legislative Effectiveness and Majority Party Size: A Test in the Indiana House," *J. of Polit.,* Nov. 1969, *31,* 1063–79.

J. T. Murphy, "Political Parties and the Porkbarrel: Party Conflict and Cooperation in House Public Works Committee Decision-Making," *Amer. Polit. Sci. Rev.,* Mar. 1974, *68,* 169–85.

W. A. Niskanen, *Bureaucracy & Representative Government,* Chicago 1971.

G. Owen, *Game Theory,* Philadelphia 1968.

C. R. Plott, "A Notion of Equilibrium and Its Possibility Under Majority Rule," *Amer. Econ. Rev.,* 1967, *5,* 787–806.

W. H. Riker, *The Theory of Political Coalitions,* New Haven 1962.

D. W. Rohde and K. A. Shepsle, "Democratic Committee Assignments in the House of Representatives: Strategic Aspects of a Social Choice Process," *Amer. Polit. Sci. Rev.,* Sept. 1973, *67,* 889–905.

A. K. Sen and P. K. Pattanaik, "Necessary and Sufficient Conditions for Rational Choice under Majority Decisions," *J. of Econ. Theory,* 1969, *1,* 178–202.

K. A. Shepsle, "On the Size of Winning Coalitions," *Amer. Polit. Sci. Rev.,* June 1974a, *67,* 505–18.

———, "Some Theoretical Results on Coalition Structures in Symmetric N-Person Games," presented at the MSSB Conf. on Uncertainty, Polit. Processes and Pub. Policy, San Diego, Aug. 1974b.

———, "A Model of the Congressional Committee Assignment Process: Constrained Maximization in an Institutional Setting," *Pub. Choice,* in press, 1975.

J. Sloss, "Stable Outcomes in Majority Voting Games," *Pub. Choice,* 1973, *15,* 19–48.

J. Von Neumann and P. Morgenstern, *Theory of Games and Economic Behavior,* New York 1944.

An Evolutionary and Institutional View of the Behavior of Public and Private Companies

By MARC J. ROBERTS*

This paper reports on an extensive study designed to explain the variations both within and among three public and three private electric utilities in their impact on environmental quality. The more general purpose is to demonstrate the value of a disaggregated model of the behavior of an organization based on the intendedly rational, but imperfect, choices of its individual members. The analysis focuses on what determines the opportunity sets, objectives, and beliefs of those members and on how the organization aggregates their choices into corporate patterns of action. Internal features of the organization—such as how tasks and authority are divided, information processed, and individuals recruited and rewarded—are thus potentially important explanatory variables, provided (as is often the case) that external pressures and constraints are not perfectly coercive.

This conceptual framework would appear to be applicable to a wide variety of public, private, and nonprofit organizations, while generating a number of the better known models of firm behavior as special cases. It is designed both to provide a framework for constructing explanations of particular past events and to give rise to testable generalizations about the behavior of organizations—particularly ones that might be useful for policy purposes. This necessarily brief paper presents only the general arguments and results, with illustrations from the case material. Since the case studies suggested much of the theoretical analysis, they do not provide strictly independent support for the propositions to be presented. (For a full account see Roberts 1975.)

I. Conceptual Framework

Real organizations do not maintain continuous, perfectly calibrated control over all their activities in order to insure "optimal" outcomes. As Herbert Simon and others (e.g., William Baumol and R. E. Quandt) have taught us, doing so is often impossible, and in any case it is generally not globally rational; the gains of perfect over imperfect decisions are typically not worth the added costs. Instead, "satisfactory" solutions are sought to a series of "problems" as these are generated by changing external circumstances, bureaucratic routines, or the poor performance of the organization in the eyes of its members.

Within the organization, the structure of tasks and authority transforms each corporate problem into a set of decision problems for each member and then combines the individual choices which result into a pattern of collective action.[1] An individual's impact on this process will depend upon, among other things, the resources and information channels he controls, his influence on others, and on how much he tries to influence each particular outcome. For the organization as a whole, the consequences of its actions will in turn depend on the responses of various external actors (competitors, buyers, regulators) and on the nature of the organization's relationship to them.

---

* Associate professor of economics, Harvard University. I am very grateful to William Capron, Richard Caves, James Duesenberry, Otto Eckstein, Richard Neustadt, and Thomas Schelling and to the participants in seminars at Harvard, Columbia, and Princeton for their comments on earlier versions of this and related papers, and to the Ford Foundation for supporting this research. The administrative and secretarial assistance of Lois Hager, the research assistance of Lane McIntosh, and the collaboration of Jeremy Bluhm on the larger study on which this paper draws have been essential to the project.

Each organization member has certain control variables such as his own time, a budget, the time of his subordinates, and so on, and certain target variables such as his immediate financial gain, his professional status, his future career prospects, etc. The various processes and practices within the organization that connect an individual's control and target variables can be thought of as a control and incentive system.[2]

For top management, the ideal control system would shape individual opportunity sets so that, in pursuing their own goals, the members of the organization make decisions that lead to the corporate actions and outcomes which top management desires. In practice, however, organizational perfection (like analytical perfection) is either infinitely or at least unattractively costly. Hence more or less crude performance indexes, output quotas, input restrictions, or other devices are often employed. The effectiveness of any control system is also limited by the fact that some individual target variables may be partly or wholly beyond the reach of the control system (e.g., the wish to exhibit technical virtuosity or a desire for world peace).

This account suggests that we can view the structure of the individual's decision problem in the manner illustrated by the flow chart in Figure I. Leaving aside some of the connecting arguments for brevity, suppose that individuals are intendedly but imperfectly rational in pursuing multiple objectives and that those objectives and beliefs about how to achieve them are influenced by past experience. Then the following four classes of variables should serve to explain how individual choices are made within organizations and then aggregated into patterns of corporate behavior.

The external environment of the organization includes the factor and product markets in which it operates, the regulatory constraints upon it, and the social and political pressures to which it is subject. These determine the consequences for the organization of its actions and hence partly define individual opportunity sets as well since individual rewards will often depend, to some extent, on what happens to the organization as a whole.

The organizational structure defines the tasks each member faces and the resources he controls to carry them out. This category includes the way authority is delegated, information communicated, and the responsibilities, size, and scope of various functional units.

The control system includes all personnel, promotion, and compensation practices that mete out rewards and punishments to members of the organization as a result of their choices and behavior. Along with the external environment and the organization's structure, it determines each member's opportunity set.

---

[1] The discussion which follows obviously owes much to James March and Simon, Alfred Chandler, and Joseph Bower. It differs from some so-called "behavioral" analyses in that it seeks to reproduce the decision problems of particular individuals and not simply the implicit logic in the firm's behavior. Compare Richard Cyert and March, chs. 7 and 8. A more complete discussion of the model here is given in Roberts (1975). A somewhat related work is O. E. Williamson.

[2] Thus, our analysis differs qualitatively from various comparative static models of cooperative and collective enterprises, which have assumed that all members confront a similar and unchanging relationship between corporate and private outcomes. See among others Jaroslav Vanek, M. Pauly and M. Redisch, and Joe S. Bain, Richard E. Caves, and Julius Margolis, pp. 276ff.

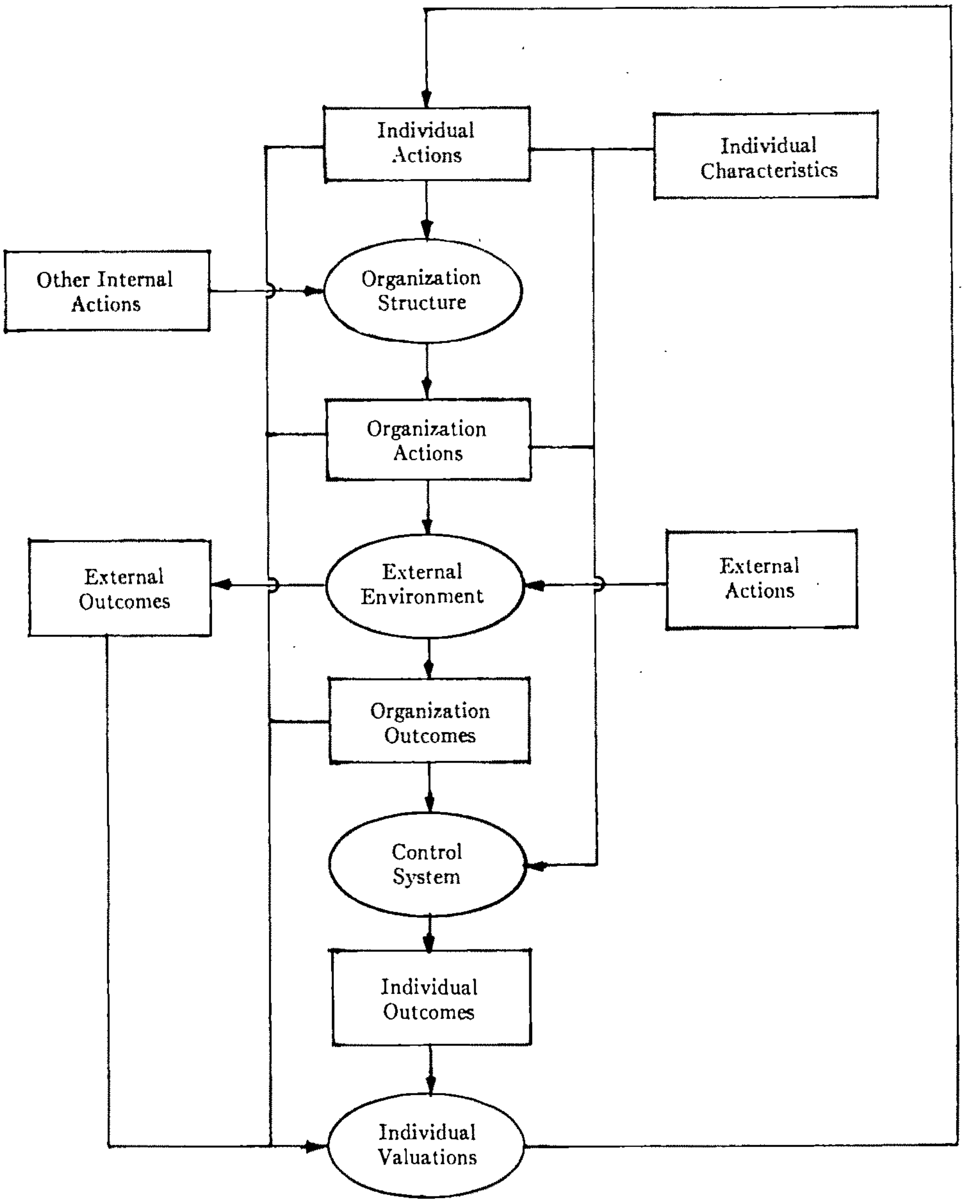


FIGURE 1: THE STRUCTURE OF AN INDIVIDUAL ORGANIZATION MEMBER'S DECISION PROBLEM

Individual objectives and beliefs, in turn, determine how the members of the organization respond to the decision problems which the organization poses to them. The beliefs held by others are also an important aspect of each individual's opportunity set. Such beliefs include personal and professional norms, beliefs which are distinctly characteristic of most organization members (which I will call the *organization's ideology*), and the plans and approach to business problems chosen by top management (which can be called the *organization's strategy*). The distribution of beliefs and values within a company obviously depends on both wider cultural processes and on its particular recruitment, socialization, and promotion practices.

Over time, the structure of an organization, its control system, and the beliefs and values of its members often change in response to changes in the external environment. The degree to which such internal features have an independent explanatory role depends on how rapidly they alter in response to such external changes and on how much discretion the organization has once such readjustments have taken place. In an imperfectly coercive environment, the internal features of an organization will be partially self-determining.[3] The current structure of an organization, its incentive system, and the beliefs and values of its personnel will all reflect its responses to past problems. Hence, an organization's immediate response to current problems, as well as its subsequent reactions if those initial efforts should prove unsuccessful, will tend to reflect its past history and development. In this sense the model is evolutionary as well as institutional although, unlike biological evolution which occurs via arbitrary mutation and natural selection, the evolution of an organization is a process of conscious, incremental self-transformation.

Two major qualifications to this characterization are in order. First, the idiosyncratic character or beliefs of particular individuals within an organization may be causally important. Such factors represent a potentially important source of unexplained (i.e., residual) variation within this framework. Second, the world outside of an organization is not completely exogenous. Instead, organizations often do act to alter what might be seen as external limits and pressures (e.g., by lobbying for new legislation or developing new technology). These additional complexities should be kept clearly in mind in the analysis of specific decisions and cases.

II. Study Design

The framework just presented suggests (as will become clear below) that the internal features of an organization will have the most impact on decisions that are uncertain, complex, and long term, and when the industry involved is imperfectly competitive. To facilitate this admittedly exploratory research, such a set of choices was selected for study: the environmental and related capital investment decisions of three public and three private electric utilities. A set of companies was chosen which exhibit much variety in their internal structures, incentive systems, and in the external situations they confront. These were: the Tennessee Valley Authority (*TVA*), the Ontario Hydroelectric Commission, the Los Angeles Department of Water and Power (*LADWP*), Pacific Gas and Electric (*PG & E*), the Southern Company, and Boston Edison.

This sample includes one public and one private company in each of three regions, which crudely controls for some variations in input prices and political context. A wide range of behavior is also represented: some of these companies have done more than formally required, while others have refused to abide by environmental regulations, sought exemptions to them, or worked strenuously for their relaxation. Furthermore, most have exhibited notably uneven performance, taking a lead nationally in some technologies or problem areas, while doing much less elsewhere.

The primary data base consists of twenty to forty interviews per company, of an hour or more, ranging from chief executives to nonsupervisory technical personnel in various functional divisions. Published and unpublished materials were reviewed and interviews with relevant outsiders were also conducted.[4] The interviews explored both the specific events that lead to certain decisions and the structure and functioning of the organization and the nature of its members more generally.

---

[3] By "imperfectly coercive" I mean that knowing simply the current external situation will not allow one to exactly predict the behavior of the organization. Instead, as in all "satisficing" models, observed outcomes depend on the history of the system.

Given the lack of published materials with the necessary detail, there was no other methodology available for establishing what exactly has occurred in each company. Interviews also provide direct information on the values and perceived incentives of various organization members, important data within our analytical framework. Intensive case studies were carried out in the hope that this would make it possible to discover to what extent particular individual actions were a response to external or internal aspects of the situation. This requires an understanding of whether there were alternative, and also functional, strategies which could have been, but were not, adopted by the organization and why events occurred as they did. Furthermore, detailed knowledge of at least a few cases makes one aware of the limitations of alternative simplified models and hypotheses and allows one to choose among them in an informed manner.

The interviews were off the record and subject to prior clearance before attribution. Obviously such data must be used skeptically, and cross-checking among various internal and external sources is essential. Yet, while some faulty memories and apparently conscious deceptions were encountered, many of those interviewed were both outspoken and critical in discussing their company's decision making. Thus it was possible to construct an overall picture of how each organization has functioned and to relate this synthesis to a detailed account of a number of specific decisions in each company. Space constraints dictate that case studies cannot be presented in detail. Instead the following section reviews the general propositions which the study suggests and briefly illustrates these from the case material.

III. Proposed Hypotheses and Findings

Since the more general results of this study were developed through the interaction of both inductive and deductive methods, their current status is that of proposed hypotheses which are nonetheless consistent with both the analytical framework and the available empirical evidence. For each of the five sets of generalizations to be discussed, we will proceed from the conceptual to the empirical, recognizing that verbal analytical arguments, like all theorizing, involve essentially a tautological restatement of the assumptions in a manner which serves to clarify their otherwise unperceived implications.

In order to construct certain hypotheses, it is useful to distinguish among various kinds of control systems. I will call the control system "weak" when, for different individual actions that lead to different organization outcomes, an individual confronts similar multivariate probability distributions with respect to firm-provided rewards and sanctions. In such cases, the individual will typically have little to gain or lose financially by various actions. Hence he must resort to other personal, professional, or organizational values in order to resolve his decision problem. The difficulty of providing a strong control system leads some organizations to rely instead on inculcating an ideology. Thus we can distinguish three types of situations.

---

[4] See Roberts (1975) for a full list of interviews and an extensive bibliography. That manuscript should be considered the source for all specific empirical statements which follow.

(1) The stronger is the control system in an organization, the more impact the values and beliefs of top management will have on the organization's choices, if top management so chooses.

(2) The weaker is the control system, and the more the organization relies instead on developing an ideology for control purposes, the more difficult it will be for top management to alter the organization's choices. In such cases, change occurs only slowly or not at all, even in the face of substantial external pressure. The gap between top management's goals and actual outcomes will depend in part on the content and flexibility of the corporate ideology and on the relationship of these to the particular aims being sought.

(3) When the control system is weak, and the organization does not possess a well-developed ideology, authority will often be delegated with little supervision since middle managers have no incentive to do otherwise. As a result, the values implicit in a decision will vary, depending upon the particular individuals who make the choice in question. Such situations are only functional when the organization is relatively free of external pressures.

To illustrate these hypotheses it is useful to begin with *TVA*, an organization with a very strong ideology and a control system that places only limited power in the hands of top management. Except for the one or two levels just below top management, promotion and compensation are decided largely in each of the seventeen divisions and offices—all of which report directly to the General Manager. With only a small staff, he and the three-man Board of Directors do not find it easy to keep informed of *TVA*'s multifarious activities, still less to supervise them closely. Policy is typically set by negotiation among the various divisions and offices. For complex historical and bureaucratic reasons, the largest and most influential of these, the Office of Power, has been committed to the notion that *TVA* should minimize power costs; and other divisions have had every reason to develop complementary, not competitive, strategies.

This internal situation became quite important when the Board of Directors voted in 1959 to achieve "clear stacks" at the Kingston power plant by installing high-efficiency electrostatic precipitators to remove fly ash. A strong ideology that had different policy implications could not be overcome by a control system that gave top management little leverage. In an effort to minimize costs, underdesign, poor construction, and inadequate maintenance all occurred. Today the precipitators work very poorly, and with state prodding, a major retrofit is being undertaken.

In contrast *PG & E* has a strong control system. Top management has much discretion in promotions and the like and participates actively in such decisions quite far down into the organization. As a result, the enthusiasm of top management for building a nuclear power plant on the coast north of San Francisco led to not one but two such attempts, despite reservations on the part of some technical personnel. The first site was rejected on geologic grounds by the Atomic Energy Commission after a multimillion dollar investment by the company. Yet only a few years later and not 100 miles away, the company made a second, equally expensive and equally unsuccessful, effort. Subordinates who were trying to please their superiors were not eager to question the wisdom of the latter's chosen course.

In contrast, since most promotion in the Los Angeles Department of Water and Power depends on Civil Service examinations, it has a relatively weak control system (and, in recent years, a weak ideology). As the proposed hypotheses suggest, the choices of this organization are not obviously consistent. Long a leader in beautifying and landscaping substations, it nonetheless has often faced vigorous community opposition over transmission line rights-of-way. In part this is because the architects who do station design see "client" concerns as more legitimate than the electrical engineers who are responsible for power-line routing. And each has some freedom to make choices in what it believes to be a professionally correct way because of the weak control system.

Our second set of hypotheses concerns the implications of this framework for cost minimization and profit maximization. Here a serious analytical difficulty arises. What standard of comparison should we use? One can argue that intendedly but imperfectly rational decision makers, by definition, never made mistakes *ex ante*. Hence, to consider a decision in error we need either more knowledge, better analytical capacity, or different objectives than the actors involved. In the following hypotheses, therefore, we employ an *ex post* standard, realizing that the test may not be "fair" in some sense, even though it is the accepted perspective for evaluating economic performance.

(4) Information and decision costs, organizational imperfections, and bad luck mean that organizations will not optimally achieve *ex post* minimum costs, maximum profits, or any other goal subscribed to by top management.[5]

(5) A weak control system, without a counterbalancing ideology, will lead to effort-reducing capital investments, overmanning, and a failure to minimize costs in routine operations. A corollary to this hypothesis and hypothesis (2) is that organizations subject to strong external pressures (e.g., survival threats) will not exhibit this structure-behavior problem.

With regard to cost minimization, most of the companies studied have had significant engineering problems. For example, one has a large coal-fired plant whose boilers are too small, and the unit cannot operate at full load for more than short periods. Another installed an innovative cooling system that has not worked properly, and that unit too cannot function at full power for extended periods. And the list could be extended easily much further. In all of these cases, cost minimization was an avowed and accepted internal goal, and extensive engineering studies were made. But because the problems were difficult, *ex post* minimization did not occur.

As to cost consciousness generally, some utilities like *PG&E* and *TVA* are very cost conscious. In the first case, the unique constitutional situation in California makes that company fearful of local referenda on expanded municipalization of its distribution network, an external constraint which has deeply influenced corporate strategy and structure. *TVA*, in contrast, has had only moderate external cost constraints,[6] but the internal mechanisms noted previously have led to similar results. *LADWP*, on the other hand, was, in the opinion of several of its own senior engineers, more "service" than cost oriented—at least until quite recently. With steadily declining costs and rates set by a City Council favorable to the department, the self-interest of organization members in large staffs, and light work loads and professional norms sanctioning technocratic excellence played a significant role in the allocation of corporate resources.

The third set of hypotheses concerns

---

[5] This also means that effectively regulated companies will not perfectly maximize net revenues by precisely adjusting their capital stock above cost-minimizing levels (the so-called Averch-Johnson effect). Since these companies all enjoyed substantial regulatory lag, however, it is not clear how much one should expect this phenomenon.

[6] *TVA* has no state rate regulation. Indeed, it acted aggressively through most of its history to force municipalities and cooperatives that resold its power to lower their rates.

what determines the strategy of an organization and the role of various groups of individuals with similar perspectives in that process. The prevalence and cohesiveness of such groups in a company depend upon how the process of recruitment and socialization and the operations of the control system serve to reinforce or break down common perspectives and incentives.

(6) The assumptions implicit in professional training, plus more or less conscious self-interest, mean that groups typically view the organization's problems in ways that place special importance on their own responsibilities and skills.

(7) Where promotion is vertical and seniority important, an individual's prospects for advancement depend heavily on the growth of the unit to which he is attached. In such circumstances, bureaucratic units will tend to become cohesive groups. In contrast, where horizontal and diagonal mobility are more frequent, broader information and different incentives decrease the cohesion of bureaucratic units.

(8) The influence of various groups will be enhanced by: the number of group members; the amount of work assigned exclusively to them; the position and personal stature of the group's senior executive; the number of group members in top management or in other bureaucratic units; and the extensiveness and coherence of prior professional training and distinct work experiences among group members.

(9) The balance of group perspectives within an organization will have more influence on its choices when the available policy options are more numerous, uncertain, apparently similar, and difficult to evaluate. In such instances (e.g., choice of a new technology), strategies that lack group advocates will seldom be adopted.

In fact, in the cases studied, groups in the sense used here were a major factor in organizational decision making. At *TVA*, where there was least horizontal mobility, bureaucratic units most clearly played such a role. In certain other companies this was less marked, in part because several had intentionally tried to rotate personnel and reorganize bureaucratic divisions in order to blur traditional distinctions and rivalries.

Group attitudes were often self-serving. Operators, who must fix plants when they fail, argue for reliability and redundancy in design. Public relations people, lawyers, financial specialists, and marketing personnel all insist on the centrality of their role and the importance of keeping "their" considerations in mind in corporate policymaking. Such stereotypes are widely accepted in the industry, as the following quotations illustrate. These come from senior engineers, one in a public and one in a private company: "He's a P.R. guy; he would give the company away if no one was looking." "When I first came here (to operations) I thought like a design engineer. Now my friends in engineering say I always sound like an operator."

Our data also confirm, in an illustrative manner, the hypotheses advanced above about the causes of a group's influence. For example, consider how the Los Angeles Department of Water and Power wound up constructing the Scattergood Unit Number 3, which the local Air Pollution Control district will not allow to be operated at over 60 percent of capacity. During and after World War II, the department underwent a conscious transformation as the Mayor brought in a new administration in an effort to remove the *LADWP* from its previous extensive involvement in local politics. One result of these changes was to eliminate many previous positions and limit the department to a small public relations group which reported directly to a single General Manager, who supervised both the water and the power systems. At the same time, the department had no legal staff of its own. Instead it used a branch of the City Attorney's office, whose members were not directly under the Chief Engineer's supervision. Hence, the management of the power system simply did not have much input from or access to specialists expert in interacting with the public or in handling regulatory processes.

Thus, when the construction permit for the plant in question was not forthcoming from the Air Pollution Control District, the department's management apparently assumed that their view that the plant was necessary simply would prevail in the end. No one in the department argued that a sophisticated analysis of the political pressures on, and the likely reaction of, the regulatory authority suggested the need for a different strategy—at least not until after construction was halted by legal action. Instead, a mechanical engineer and a busy top executive handled the matter, with very undesirable results from the department's viewpoint. Such events would be very unlikely, in contrast, in an organization like *PG & E*, a company very conscious of public opinion and political constraints, with several lawyers in key general management positions and very large in-house public affairs and legal departments deeply involved in most significant decisions.

Particular groups with particular interests have also been responsible for most of these organizations' ventures into new technologies. At the Southern Company, the Research Department entrepreneured an extensive program of research on solvent-refined coal. At *TVA*, the Air Quality Branch has taken the lead in developing operational control methods for meeting ambient air quality standards by varying power plant operations. A group of mechanical engineers at Boston Edison took the initiative and did the work that made possible that company's extensive involvement in stack gas-scrubbing experiments. And the transmission engineers who lead the *LADWP* were crucial to that organization's decision to construct a pioneering 800 kilovolt D.C. intertie to the Bonneville Power Authority. In all these cases, the companies have been leaders nationally in these particular areas, while not being notably adventurous in other respects.

The fourth set of hypotheses concerns the influence of the political system, which ultimately has great power over both the structure and the external situation of most organizations in the society. Note, however, that in practice the legal limits it can impose are not inviolable constraints. Rather they confront the firm with a set of multivariate probability distributions over various costs and consequences (fines, delay, etc.) contingent on its own actions.

(10) Because the categories "public" and "private" include organizations that exhibit a wide (and overlapping) set of internal features and external circumstances, that distinction contains little predictive information about behavioral differences.

(11) To the extent that legal arrangements raise the transaction costs that political actors face in attempting to control an organization, the organization will confront relatively less restrictive external limits.

> If the organization's budget requires routine political approval, the costs of political control tend to be lower and the extent of such control correspondingly higher.
>
> If political control requires the coordinated action of more than one political entity, the costs of such control tend to be higher and its extent correspondingly lower.
>
> If different individuals and political actors face different costs and benefits from altering the policy of an organization, those who face the lower costs and stand to enjoy the higher benefits will be

> more favored by organization's decisions. Such variations in costs and benefits will typically develop when the geographic scope of an organization's activities is smaller or larger than that of the political entity to which it is nominally responsible.

On the public-private distinction, our studies show that some public and some private companies are cost conscious and others are less so. Some in each category have made great efforts in the environmental area while others have done relatively little. The critical role of government control over access to capital is also apparent in the cases. An organization like Ontario Hydro, whose budget is routinely reviewed and approved by the Provincial Cabinet, is far more vulnerable to (and hence responsive to) political pressures than *TVA* or *LADWP*, which are financed from retained earnings and revenue bonds and whose budgets do not require anyone else's approval.

The cases also suggest that when a company is smaller than the scope of the relevant political system, as *TVA* is versus the national government, it becomes much less likely that the political entities in question will mobilize to control the organization's policy. In such instances, after all, that policy is of interest to only a minority of the government's citizens. On the other hand, *LADWP* illustrates the hypothesized tendency for an organization to treat those who face high costs in influencing it differently than those who face low costs. With the urging of the County Air Pollution Control District, the department took the view in the late 1960's that the environmental problems caused by generating power for Los Angeles could and should be solved by constructing new facilities in other areas—whose citizens voted in neither city nor county elections.[7] This part of the argument is especially significant in light of the suggestion often advanced in recent years for the creation of special regional agencies to deal with various environmental purposes. Unless these are carefully designed, they would appear to be very costly to control in terms of political resources, and hence less likely to be so controlled (cf. the New York Port Authority).

Finally, we can offer some hypotheses about the evolutionary process by which organizations respond to changing external circumstances. To some extent, such changes reflect whatever capacity top management possesses to purposefully alter the structure, incentives, and personnel of the organization, although the results of such actions are not always fully anticipated. Yet, the goals and strategies that shape such management actions are not themselves entirely exogenous; top managers, after all, are typically chosen by their predecessors or by others within the organization.

(12) Since individuals generally prefer to be identified with socially acceptable organizations, changing social attitudes will tend to result in some changes in an organization's behavior, provided that other pressures or beliefs do not have different implications.

> Because newer and younger members often have more "modern" views than those they replace, more rapid personnel turnover facilitates changes in an organization's behavior in currently socially sanctioned directions.
>
> For top management to alter a widely held strategy in an organization with a strong ideology, it may be necessary to advance new individuals, especially outsiders, to key positions.

(13) Changes in an organization's struc-

---

[7] The department has been a major participant in coal-fired power plants in Arizona and Nevada and was a central advocate of the unsuccessful Bridge Canyon proposal, which involved a dam on the Colorado River that would have backed water up into Grand Canyon.

ture undertaken to facilitate the accomplishment of chosen policies not infrequently lead to unanticipated changes in those policies.

(14) When an organization's chosen strategy appears to be unsatisfactory, the initial response often is to make an effort to change one of the binding external constraints.

The detailed studies have developed much interesting material on these last issues. In almost every company some members, especially younger ones, were willing to criticize their own superiors for not doing more about air and water pollution. Similarly, time and again, a change in management has been associated with a change in policy. In Los Angeles, for example, several City Hall observers commented on the more flexible and publicly sensitive approach of the new management of the Department of Water and Power—an observation made as well by a number of those inside the organization.

The impact of changing social attitudes is very evident. Most older executives assert that their views, or the organization's policy, or both, have changed in recent years. The typical position is that the only real issue is how much to spend on environmental protection to correctly balance the costs against benefits. Ten years ago these men, and by their own accounts their predecessors, would have taken a noticeably different position.

Changes in policy obviously also arise because chief executives have consciously altered the structure and personnel of their organization in order to alter its behavior. But their ability to do so is often not unlimited. In *TVA*, for example, most individuals at all supervisory levels share a commitment to producing power at the lowest possible costs as the dominant policy goal. Perhaps in response, in order to obtain more policy flexibility and personal control, the current General Manager has made at least three senior appointments in recent years from outside the Authority.

There are also cases where changes undertaken for one reason have had unanticipated effects. One company created a Research and Environment Division that soon moved energetically to expand the corporation's environmental research and development program. And, if not always successful, there is reason to believe that it has succeeded more than some foresaw when it was first established.[8]

The cases also show organizations responding to "problems" by trying to change nominally external pressures and limits. For example, both the Southern Company and *LADWP* pleaded poverty and halted work on major projects during rate disputes. Intentionally or not, in both cases this provoked political efforts in favor of rate increases by the construction unions whose members were thereby unemployed. Just recently *TVA* lobbied a bill through Congress to have the Treasury in effect pay for its pollution control investments (only to have it vetoed). The willingness of organizations to take various risks and to attempt to relax one or another constraint depends in part on the details of corporate policy formation and the balance among group perspectives as discussed previously.

IV. Conclusions

The preceding discussion implies that there may be no simple explanation of the inter- and intra-corporate behavior patterns that we are interested in. *PG&E* has "clear stacks" and *TVA* does not. On examination this difference turns out to be a function of a complicated set of differences in history, fuel prices, external pressures, the strength of internal control mechanisms, and the balance of group perspectives within each organization.

---

[8] In this company, budgetary procedures for research have become significantly more formalized in the last year.

The proposed hypotheses admittedly do not impose highly restrictive limits on the behavior of organizations. Like consumer theory in general, one cannot get detailed results from general assumptions. The argument has assumed little more than that people are intendedly rational, that decision making and information are expensive, and that tastes and beliefs are influenced by experience. To get more specific implications, one needs to postulate more specific conditions. For example, hypotheses (6) and (8) together might be specialized as follows: "In electric utilities, power plant operators value plant reliability. Hence, the forced outage rate of a system's facilities will be inversely related to the seniority and influence of the company's chief operating official and to the percentage of the organization's technical personnel who were operators at the time the plant was designed."[9] That generalization, and others like it, I hope to subject to statistical confirmation in future research efforts.

On the other hand, not all that can be learned about organizations and their functioning can be summarized in a set of formal hypotheses. The design of organizations for policy purposes, like, say, the design of sailing boats, will retain an important component of nonspecifiable craft skill for the foreseeable future (Roberts 1974). That does not mean, of course, that we should not try to systematically write down as much as we can about the orderly features of the relevant phenomena. But I believe that there is a value in being familiar with the details of particular situations, for policy as well as historical purposes, hence, the detailed manuscript which includes the full case studies, which is in preparation.

Finally, I contend that this study has been fruitful enough to show that there is something to be learned from disaggregating economic entities in order to better conceptualize their behavior. No doubt simple profit maximization explains a good deal in some cases. But it explains much less in others. And shouldn't we want to be able to account for such variations? Modern chemists explain the behavior of a substance by reference to its molecular and atomic substructure. Perhaps economists have been a bit too optimistic and, like medieval chemists, have viewed the world in terms of categories that are too few and too aggregated. If the disaggregation and detail of the evolutionary institutional viewpoint can improve our understanding, generate new hypotheses, and reduce the margin of unexplained residual variance, it would seem to be worthy of further attention, especially in those cases like public enterprises, where no "black box" model readily suggests itself.

REFERENCES

J. S. Bain, R. E. Caves, and J. Margolis, *Northern California's Water Industry*, Baltimore 1966.

W. J. Baumol and R. E. Quandt, "Rules of Thumb and Optimally Imperfect Decisions," *Am. Econ. Rev.*, Mar. 1964, *54*, 23–46.

J. L. Bower, *Managing the Resource Allocation Process*, Boston 1970.

A. D. Chandler, Jr., *Strategy and Structure*, Cambridge, Mass., 1962.

R. M. Cyert and J. G. March, *A Behavioral Theory of the Firm*, Englewood Cliffs 1963.

J. G. March and H. A. Simon, *Organizations*, New York 1958.

---

[9] For example, Ontario Hydro, which has had relatively high outage rates on some older units, at the time had no central unit dealing with steam power plant operations. Instead, operators reported to regional supervisors who had multiple responsibilities for line maintenance, service, sales, etc.

M. Pauly and M. Redisch, "The Not-for-Profit Hospital as a Physician's Cooperative," *Am. Econ. Rev.*, Mar. 1973, *63*, 87–99.

M. Roberts, "On the Nature and Condition of Social Science," *Daedalus*, Summer 1974, 47–64.

———, *The Evolution of Environmental Decisions*, book manuscript, 1975.

H. Simon, "A Behavioral Model of Rational Choice," *Quart. J. Econ.*, Feb. 1955, *69*, 99–118.

J. Vanek, *The General Theory of Labor-Managed Market Economies*, Ithaca 1970.

O. E. Williams, *Corporate Control and Business Behavior*, Englewood Cliffs 1970.

THE PRINCIPLES COURSE: WHAT SHOULD BE IN IT AND WHERE SHOULD IT BE GOING?

Some Observations on the Learning of Economics

By KENNETH E. BOULDING*

We probably pay too much attention to the teaching of economics and not enough to the learning of it, that is, to those processes of reorganization in the cognitive structure or image which result from being exposed to the body of public knowledge which economists share. All learning processes involve the reconstruction of images. This sometimes involves fairly drastic reorganization change in existing images; sometimes it merely involves addition to the existing structure. Thus, when we learn that Sacramento is the capital of California, this simply adds to our geographical knowledge of the United States without any very fundamental change in the structure. When we learn, however, that the shortest way from California to Iran is over the North Pole, if we have previously structured our world on Mercator's Projection, this is quite a shock and will create a considerable reorganization of our geographical image. The learning process, therefore, depends a great deal on where people start from, that is, what images of the system they have in their minds at the beginning of the process; and it depends also on the creation of certain shocks or inconsistencies, what Leon Festinger has called "cognitive dissonances," which force a reorganization of the structure of these initial images.

In discussing the teaching of economics, we have usually been singularly insensitive to the problem of what images of the economic system people start with. If we don't know what it is we are modifying, we may not be successful in modifying it. In a sense the object of formal education is to move images in the minds of the students from private and idiosyncratic images toward publicly held images. This is not to say that the public image is always right. Indeed, change in the public image always arises because some private, idiosyncratic image fails to correspond to it, although it corresponds more to the truth. Nevertheless, for most purposes we assume that the teacher's image is "right" and that the student's initial image is "wrong" and that the business of education is to change the image in the mind of the student in the direction of the image in the mind of the teacher.

If we look at the learning process in this way, it might occasion a certain change in our methods of teaching. The traditional method is to create a large flood of information input into the student's ear, and perhaps into his eye, in the hope that this input will create a change in the student's image in the direction of that held by the teacher. Most of the feedback to the teacher, however, comes from his own internal perceptions as to whether the information stream which he is emitting is consonant with the image which he holds.

---

* University of Colorado.

He gets very little feedback from the student and is often quite unaware that the information stream which he is transmitting may have a completely different effect on the image of the student from what it is supposed to have. It may, of course, go in one ear and out the other; it may simply be filtered out; or it may be filtered in such a way as to create transformations in the student's image which do not correspond to the intentions of the teacher.

Once we concentrate on the learning process rather than on the teaching process, we tend to start off with creating in the mind of the teacher an image of what is the initial state of the mind of the student. The process should begin, therefore, with some sort of inventory or examination which will improve both the student's and the teacher's perception of the student's present image. This could include a vocabulary test to investigate the student's image of the meaning of the major terms in the field. It should also include a test of quantitative estimates to get at a student's image of the quantitative proportions of the economy. I find even with graduate students, for instance, that they grossly overestimate the proportion of national income going into profits. Another useful inventory would probe what the student sees as the major problem of the subject, that is, what questions does he want this course of study to enable him to answer. A very good exercise indeed might be to make the student at the outset write a set of examination questions so that the student is in effect setting an examination for the teacher.

Once these initial inventories are in, the student should be faced by a cognitive dissonance, that is, by whatever contrast there may be between his own image of the world as revealed in the inventory and the teacher's image, which one hopes is part of the public image. A course of directed reading might follow, designed to enrich the student's vocabulary, acquaint him with concepts, and generally modify his image in the direction of that of the teacher or of the profession. In the natural sciences, it is possible to give the student laboratory experience of the "real world," although this is often pretty contrived, and what the student is often really doing in the laboratory is trying to find the answer that will please the teacher, just as he is in class. In the social sciences, laboratory experience is much more difficult. Nevertheless, the presentation of historical and factual materials and then setting the student to the task of interpreting these materials according to his own images of the world may have all the impacts of the laboratory. Statistical time charts are of particular value in this connection; so are maps, distributions, and expressions of quantitative data.

The hardest part is, of course, the image of relationships or models. Here the student may have extremely vague ideas before he begins to study. Hence, what is happening is not so much the reorganization of existing inaccurate images, but the creation of new kinds of images altogether. How this is done is a bit of a mystery. The response of a student depends a great deal on his "readiness," but this is very hard to discover. I have had students who have just had an experience of enlightenment come to me and say, "Why didn't you tell me that six months ago?" to which I usually reply, "I did, but you weren't ready for it." It may be that the only answer to the readiness problem is a good deal of redundancy in the teaching process. It is not so much that what I tell you three times is true, as what I tell you three times may be believed on the third try. It may well be, therefore, that the presentation of material can be too orderly in a field which is a "round globe of knowledge," like economics, where the understanding of one part depends very much

Some Comments on the Principles Course

By JOHN G. GURLEY*

Quite aside from what *should* be taught in our principles course, what is in fact taught is a system of economic ideas which come from social practice—from our efforts over time to develop our productive forces and from the class struggles that have changed our social relations. These struggles with nature and with one another govern pretty much what we think. This suggests that one way to teach elementary economics is to relate the development of economic thought to the development of society's economic base. In this way, students would be given a feeling for the relation of theory to practice, some sense of history, and the knowledge that systems of ideas do not drop out of the sky. This is what Joan Robinson and John Eatwell have done in their recent text, *An Introduction to Modern Economics.* E. L. Hunt and H. Sherman in their text, *Economics*, also follow this procedure. Such an approach is necessarily radical, for it cannot help but expose the extent to which economic theories are rationalizations of advantages held by ruling classes.

Since our own social practice takes place within the capitalist mode of production, in which the dominant social relations are bourgeois and the leading class is the capitalist one, our principles course mainly reflects and supports this type of social arrangement, so that the valuable lesson taught by the above authors is missing. Instead, capitalism is presented as a generally harmonious, equilibrating, efficient, and equitable system—at least potentially so—as a system with problems which are largely solvable with intelligence and good will and, if history is mentioned at all, as a system that represents the zenith of an historical process. Profits and other property income are justified and raised to the same moral plane as wages; and the real, underlying exploitative relations of capitalist society are hidden under a layer of exchange equalities and contract freedoms.

However, in a capitalist society, when the working class has come of age and its own ideologists have gained sufficient understanding of how the system works, then another complex of economic ideas appears, a complex which reflects the position and interests of the working class, reveals the exploitative relations and contradictions of capitalism, and shows its salient features and its transitory nature.

Therefore, in a class society such as ours, there will necessarily be two major—and no doubt several minor—ideological reflections of that class structure. This is why, although the dominant economic theory is bourgeois, there also exist principles texts expounding radical, institutional, romantic, and other approaches. At this level of discussion, the question of what "should" be taught is primarily a matter of which class to support and secondarily a matter of form.

For a long time, up to World War I, the capitalist mode of production destroyed all barriers in its headlong drive for hegemony around the world. This overwhelming material force of commodities and military weapons was reflected in the intellectual dominance of bourgeois thought in economics and other disciplines. In the past half century, however, at first slowly and lately more rapidly, international capitalism has been weakened. Some of the forces responsible for this have been the two world wars between the newly rising and the already established capitalist powers; the expansion of communism, which has now swept across a third of the globe; the rise of national liberation and ultranationalist movements throughout much of the underdeveloped world; the growing control exerted by third world countries over the raw materials and natural resources so badly needed by the capitalist nations; and the dwindling supplies of cheap labor available to the leading capitalist countries, which have resulted in the strengthening of labor relative to capital. In the past few decades, *U.S.* capitalism in particular has also been hit by intense competition from other capitalist powers. These adverse winds have tended to erode profit opportunities of *U.S.* capitalist enterprises, which they, supported by the State, have attempted to restore by exerting additional pressures on the working classes.

---

* Professor of economics, Stanford University.

Such dramatic changes in fortune in the material world opened the way for increasingly successful intellectual forays against the growing disarray of bourgeois thought. In economics, this has taken the form of a renewed challenge by Marxism, radicalism, and alternative approaches to conventional, neoclassical analysis. Bourgeois economists have been busily engaged in fending off these challenges and in trying to patch up the conventional analysis faster than new disasters and puzzles appear. Accordingly, we now find in our principles texts, not only signs of apprehension about existing theory, but also new sections on tradeoffs between unemployment and inflation and between growth and environmental decay, the woes of the *U.S.* dollar, the impact of multinational corporations, and the continuing failures of capitalism to alleviate poverty and misery in much of the world, including in our own backyards. Even though the authors put the best face they can on these problems, it is quite understandable that, in their search for further guidance or to discredit obvious alternatives, they should also add chapters on Marxism, comparative economic systems, and planning and controls. It is likewise no surprise to find some of them giving more attention to the history of economic thought as their own doubts increase about the continuing viability of conventional doctrine. When one suspects the child, one investigates its parents.

Thus, it has not been a matter of freely searching for what seems to be ideally right to put in our principles texts, but rather one of trying to understand and adjust to the powerful historical forces that so heavily bear down on us. Since these forces have favored economists who attack capitalism, conventional economists have been put more and more on the defensive, and the principles texts clearly show this. In the end, the Vietcong will probably have more to do with changing our principles course than all the roundtable debates such as this one from now to kingdom come.

While conventional principles texts have certainly changed in response to real movements in the world, the changes have not been uniformly for the better. Some authors, on the defensive, have simply reiterated, only in louder and louder tones, that capitalism's mounting problems are solvable within the system itself, by making the system work better. Others treat such topics as Marxism, socialism, and capitalism (as a system) most inadequately. Thus, we find in the principles texts that capitalism is often defined so as to hide its class nature. One author of a popular text even believes that the word "capitalism" is too emotionally charged to be safely used. He and other authors often put the word in quotation marks, as though it is being gingerly and warily held up by its corners for fear of contaminating its holders. When it comes to Marxian economics, conventional authors make one mistake after another and appear to have the weirdest notions of what the subject is all about. Some make a real mess of the whole thing. Even Paul Samuelson, who has by far the best discussion of Karl Marx among the textbook authors I have read, presents Marx incorrectly at times, superficially at other times, sometimes shabbily, and often with trickery. Few authors seem aware that there have been other Marxists than Marx—that, over the past century, a large and rich Marxian literature has been fashioned which is valuable for understanding modern-day Marxian economics. Marxism is not only generally emasculated by these authors, but it is also treated in such isolation from the rest of the material that it appears to be some new, exotic "field" in economics, rather than what it is, a mode of thought that challenges every single aspect of conventional economic theory—the ideological counterpart of the working classes' challenge to every single aspect of the monopoly of private property ownership.

I do not think, therefore, that Marxian economics can be properly presented in a bourgeois textbook. It should and will make its appearance increasingly in radical textbooks and in books of readings which will compete with the conventional offerings. The outcome of this intellectual clash will be determined by the outcome of the many struggles between the capitalist and working classes, which are now being waged on every part of the globe, for the ownership and control of the world's means of production and natural resources. Marxists believe, to put it bluntly, that marginal productivity theory will disappear only with the bourgeoisie.

Finally, with regard to what I have called the secondary consideration (but, still, one of great importance)—i.e., the form in which any set of economic ideas is taught to beginning economics students—the ideal way would be some combination of theory and practice, in which the students, while learning theory, directly investigate problems in the real world to which their theory may be applied. But since existing arrangements hardly encourage this, the closest and next best method is to teach only a small amount of theory to permit sufficient time for students to apply it to case studies as they go along. In this I agree with what Rendigs Fels said last year, when he advocated less stress on theory that is important mainly to the professional economists (and which, in any case, I would add, is mostly empty), more stress on a few basic macro and micro tools of analysis, and more emphasis on the use of these tools to solve actual problems. Students now have so much theory thrown at them—especially right now when theories are proliferating to cover up failures of previous theories—and so little time to use the tools that they very soon forget just about everything they have heard or read. G. L. Bach, Richard Attiyeh, Keith Lumsden, Fels, and others surely are on the right track in designing interesting and challenging case studies for beginning economics students to work on. This is the direction the principles course should be going.

New Approaches to Teaching the Principles Course

By W. LEE HANSEN*

Had any of us been asked to serve on a panel like this ten years ago, I assume we would have been perplexed about what to say, and this would not have reflected the fact that all of us were then ten years younger. At that time there seemed to be little dispute about the content and direction of the principles course. A substantial consensus existed within the profession as to what economics was all about. This was reflected in the second, third, or fourth editions of those superbly produced, massive forty-five-chapter principles texts, which, replete with color diagrams and accompanying readings and workbooks, covered most everything economics had to offer. This made the role of the instructor an easy one, for after selecting a text, the book did the instructing. Everyone seemed satisfied. Students weren't complaining. And faculty members had ample time to turn to other things, among them their research, the building of graduate programs, and faculty recruitment.

But today I expect those of us on the panel have more thoughts racing through our heads than any of us can present to you. What does seem clear, however, is that the principles course is in a state of transition. There is vastly greater variety in the nature and content of principles courses, the teaching materials employed, and the teaching approaches instructors are developing. These changes reflect a response to a variety of largely unanticipated developments over the past decade. The emergence of new issues and problems—the Vietnam war, poverty, environment, and the like, coupled with the challenge posed by radical economics—has called for new, or at least different, approaches than those customarily used in principles courses. Closely related is the shift in research interests of faculty who perceived a growing gap between what was traditionally taught and the things of most interest and concern to them. At the same time student pressures for better teaching, greater "relevance," and other academic reforms sensitized faculty to the need for rethinking their teaching of the principles course, both with respect to what they were teaching and how they were teaching it. In addition, the small but rapidly growing body of research on economics instruction indicated that the effectiveness of the standard principles course left much to be desired. Then during the past couple of years intensive competition for student enrollments (because of their implications for faculty staffing) has emphasized the need to make economics courses more attractive to our students. And most recently, the confusion both inside and outside the profession about how to cope with current policy issues, particularly inflation and unemployment, indicates that much remains to be done to develop the required degree of economic literacy.

What kinds of responses do we observe to these diverse forces? First, a small but growing number of faculty members are revamping their principles courses. They are shifting away from the long-standing practice of trying to survey the entire field of economics. Instead, the focus has been on narrowing the scope of the course to something far more manageable, in the hope that longer lasting learning will be the result. Concern over the ever expanding scope of the principles course was voiced as long ago as 1950 in the American Economic Association's report on *The Teaching of Undergraduate Economics.*

---

* University of Wisconsin-Madison.

Second, rather than preparing students for upper-division courses most of them never go on to take anyway, a more applied emphasis is emerging. It appears to be motivated by the question: How can economics be taught so that students by the end of the course can understand the kinds of economic problems they will be reading about and involved in throughout their adult lives? We see this changed emphasis reflected in the breakup of the principles books into micro and macro sections, in the development of a wide array of new smaller "problems" texts, and in the welcome proliferation of paperbacks and reprints on special topics. Taken together they permit the creation of one's own textbook through a skillful combination of available materials, supplemented by articles instructors choose to reproduce out of daily newspapers.

We also find an impressive collection of completely new paperback texts which capture the essence of various new experimental and innovative principles courses, many of which have been supported by the Joint Council on Economic Education. Let me list a few of these courses. We have the "case approach" developed by Rendigs Fels and others at Vanderbilt, the "current issues approach" reflected in the Oklahoma State course created by Richard Leftwich and Ansel Sharp, the "core concepts, problems approach" followed by Phillip Saunders at Indiana and also by our chairperson, G. L. Bach at Stanford, and the "social systems approach" being developed by fellow panelist Kenneth Boulding at Colorado. In addition, similar developments are taking place elsewhere; included is the emergence of radical economics courses at the principles level. We also see a few cases in which individuals are putting their own specialized stamp on a course, as exemplified by John Kuhlman's course at Missouri. What I find most heartening about all of this is the heavy involvement of individual faculty members across the country in developing new approaches that most closely fit their conception of what economics is all about.

Another noteworthy change is the experimentation going on with alternative teaching techniques. Elisabeth Allison at Harvard has developed a Keller-type, self-paced approach; Donald Paden at Illinois is working with a *PLATO*, computer-based teaching system; Allen Kelley at Duke operates an ambitious, diversified instructional program centered around his *TIPS* system; and so on. Obviously, the costs of course development, both to individual faculty and their supporting institutions, are substantial, but only by making such investments can new, more effective instructional systems be identified and developed.

Closely related is the emphasis more and more economists now give to the evaluation of their instructional efforts. Increasingly, we find them applying the research tools of their trade to measuring how the various elements of instruction and student behaviors and abilities interact to produce greater (we hope) student learning and interest in economics. The results of this work are finding their way into print, in the flourishing *Journal of Economic Education.* The establishment of this journal, by the way, represents a triumph in economic thinking, for it has cleverly helped convert teaching and research into complementary rather than competing activities that vie for faculty time and effort.

Last are the accelerating efforts to give more attention to pedagogy in the training of graduate students who anticipate teaching careers. The economics profession has long recognized its failure to do an adequate training job; this was made abundantly clear in the Association's now almost-forgotten 1953 report on *Graduate Education in Economics*. In an effort to prepare aspiring faculty members for their most important professional activity, that of teaching, the Joint Council, in collaboration with a number of major Ph.D.-granting departments, has inaugurated a teacher training program. This five-year project cannot help but heighten the awareness and interest of younger economists in the all-important work of teaching, especially the teaching of the principles course.

The length and scope of this list of recent developments strike me as impressive given the relatively low weight we assign to the teaching of freshmen and sophomores as compared to economics majors and graduate students. But impressive as this list may be, we should not delude ourselves that shifts in content and emphasis, innovative teaching materials, and new pedagogical techniques will themselves revolutionize economics instruction. Rather, whatever success is achieved in producing improved teaching and learning will be the result of greater attention by faculty members to such things as the development of course objectives, better organization and preparation, and greater demonstrated interest in what they are trying to teach their students.

Some of you may view these efforts as representing patchwork when major change is needed. Perhaps this is true. On the other hand, there is no best way to teach economics, however we define it. And there is no single definition of economics that will satisfy all of us. This argues for diversity, as instructors seek to find and develop approaches and techniques that add to their own intellectual excitement and give them a greater sense of craftsmanship. Of course, this assumes that instructors will be challenging their students to learn and in the process will be preparing them to understand the economic issues and problems they will encounter in their adult lives. It also argues for greater attention to disseminating information about new developments in the teaching of economics so that we can all learn more about how other people approach their teaching.

We should also remember that the vast bulk of our principles students will never take another economics course; indeed, many of them now enroll only because they are required to do so. Thus, the task is one of, and I don't like this phrase but can find no other, "turning people on" to economics, not in the sole hope of gaining more majors but rather of imparting to those students who do take economics an understanding that can serve them effectively long after many of the specifics they learned have been forgotten.

This poses a real challenge that each individual and institution will have to meet in its own way. One approach we have tried at Wisconsin in our one-semester principles course (we also have a two-semester sequence) is to adopt a modular format. A number of different faculty participate in the teaching of this single course, with its three lecture sections enrolling a total of over a thousand students. The course begins with a common core of four weeks of applied micro taught by one faculty member, followed by three weeks of macro taught by another faculty member. After that students can choose from a variety of three-week topical modules, taught by still other faculty. The idea of the topical modules is to apply the core material to topics of widespread interest and concern. Thus, we have modules on environmental economics, the role of

AMERICAN ECONOMIC ASSOCIATION

PROCEEDINGS
OF THE
EIGHTY-SEVENTH
ANNUAL
MEETING

SAN FRANCISCO, CALIFORNIA
DECEMBER 28-30, 1974

MINUTES OF THE ANNUAL MEETING
SAN FRANCISCO, CALIFORNIA
DECEMBER 29, 1974

The Eighty-seventh Annual Meeting of the American Economic Association was called to order by President Walter W. Heller at 4:05 p.m., December 29, 1974, in the San Francisco Hilton Hotel.

It was VOTED to approve the minutes of the meeting of December 29, 1973, as published in the *American Economic Review*.

The Secretary (Rendigs Fels) presented the report of the Committee on Elections and the certification of the new officers for 1975 as follows:

In accordance with the bylaws on election procedure, I hereby certify the results of the recent balloting and present the results of the Nominating Committee, the Electoral College, and the Committee on Elections.

The Nominating Committee, consisting of James Tobin, Chairperson, Allen C. Kelley, George L. Perry, Roy Radner, Christopher A. Sims, Alice Vandermeulen, and Charles Z. Wilson, submitted the nominations listed below for Vice-President and member of the Executive Committee. The Electoral College, consisting of the Nominating Committee and the Executive Committee meeting together, selected the nominee for President-elect. No petitions were received nominating additional candidates.

President-elect

Franco Modigliani

Vice-Presidents	*Executive Committee*
Martin Bronfenbrenner	Carolyn Shaw Bell
Leonid Hurwicz	Jack Hirshleifer
James N. Morgan	Dale W. Jorgenson
Charles L. Schultze	Burton A. Weisbrod

The Secretary prepared biographical sketches of the candidates and distributed ballots in September. The Committee on Elections, consisting of Wilfrid L. David, Chairperson, Malcolm Getz, and Rendigs Fels, *ex officio*, canvassed the ballots and filed the results with the Secretary. From the Report of the Committee on Elections, I have the following information:

Number of envelopes without name for identification	180
Number of envelopes received too late	91
Number of defective ballots	16
Number of legal ballots	5,396
	5,683

On the basis of the canvass of the votes, I certify that the following persons have been duly elected to the respective offices:

President-elect (for a term of one year)
Franco Modigliani
Vice-Presidents (for a term of one year)
Martin Bronfenbrenner
Charles L. Schultze
Members of the Executive Committee (for a term of three years)
Carolyn Shaw Bell
Burton A. Weisbrod

The Secretary-Treasurer and the Managing Editors of the *American Economic Review* (George H. Borts) and the *Journal of Economic Literature* (Mark Perlman) briefly discussed written reports which had been reproduced and distributed in advance of the meeting. The Director of *Job Openings for Economists* (C. Elton Hinshaw) gave an oral report. The Chairperson of the Committee on the Status of Women in the Economics Profession (Barbara B. Reagan) summarized the written report scheduled for publication in the *Proceedings*. The President expressed gratitude to the Committee for its outstanding work.

The Secretary presented the following

resolutions which were adopted by acclamation:

> BE IT RESOLVED that this meeting record a special vote of thanks to the members of the 1974 Local Arrangements Committee headed by Richard Holton for their hard work and efficient management of the meetings of the Allied Social Science Associations.
>
> BE IT RESOLVED that this meeting commend Robert Aaron Gordon for planning a program of such unusual interest as to compete successfully against the attractions of San Francisco.

After expressing his thanks to officers and employees of the Association for their assistance during his term of office, President Heller introduced the new President, Robert Aaron Gordon, who took the chair. He announced that if the business of the meeting was not completed by 6:15 p.m. it would be adjourned until 9:30 p.m.

Under a procedural rule adopted by the Executive Committee on March 8, 1974, and communicated to the members in connection with the distribution of ballots for the 1974 election, resolutions proposed by members had to be submitted to the Secretary a month in advance of the meeting for reproduction with the signatures of the mover and the seconder. Five such resolutions had been submitted and distributed before and at the meeting.

The new President called for Theodore P. Scheinman to discuss the resolution he had submitted calling for establishment of a nonpublications membership. Since Scheinman was not present, consideration of his resolution was deferred till later in the meeting.

Four resolutions had been proposed by William H. Behn and seconded by Henry M. Levin. The President announced that in the opinion of the Association's Counsel, as well as his own, the fourth of these resolutions urging economics departments to hire Marxian economists was prohibited by the Charter of the Association and therefore out of order.[1] There was no appeal to the ruling.

The President then called upon Behn to discuss the first of the resolutions he had submitted, which read:

> We move that the American Economic Association adopt the following resolutions:
>
> WHEREAS economics students may be subject to political discrimination with regard to employment, fellowships and research grants, and being that Federal laws provide no protection against political discrimination, therefore:
>
> BE IT RESOLVED that the American Economic Association urges economics departments and faculty to follow the

---

[1] The full text of the proposed resolution follows:

We move that the American Economic Association adopt the following resolution:

WHEREAS the academic community prides itself on the free exchange of ideas, and

WHEREAS the Marxian economic paradigm has a long intellectual history, and continued study and research in this paradigm is essential to a healthy critical approach to economics, and

WHEREAS there exist circular conditions in the United States, supported by political discrimination, which inhibit the study and conduct of research of Marxian economics [i.e., the absence of permanent (tenure track) faculty teaching and conducting research in the Marxian economic paradigm—few course offerings in Marxian economics—no student research or dissertations—few Marxist scholars—the absence of permanent faculty teaching and conducting research in the Marxian economic paradigm], and

WHEREAS the appointment of visiting and temporary (one- or two-year) nontenure track faculty are inadequate measures to reverse this political discrimination in hiring and the circularity of conditions inhibiting the study of Marxian economics, therefore:

BE IT RESOLVED that the American Economic Association urges economics departments to take immediate measures to attract and hire permanent (tenure track) economists working in the Marxian economic paradigm.

Proposed by: William H. Behn, regular member of the *AEA* (address: P.O. Box 6887, Stanford, Calif. 94305)
Seconded by: Henry M. Levin, regular member of the *AEA* (address: School of Education, Stanford University, Stanford, California 94305)
November 25, 1974

> guidelines outlined below in relation to student departmental files and job search:
>
> a. Student files, including letters of recommendation, should be open to a faculty member designated by the student to examine the file for material of a political discriminatory nature.
>
> b. Departments in their placement efforts should not release information on the extradepartmental political activities of any student, nor on the intradepartmental activities that might convey information about the student's political views.
>
> c. Faculty members should refuse to respond to inquiries concerning a student's political views that arise with regard to employment, fellowships, research grants, and other opportunities for professional advancement.
>
> d. These guidelines shall be published in each issue of the American Economic Association publication *Job Openings for Economists* (*JOE*).
>
> Proposed by: William H. Behn, regular member of *AEA* (address: P.O. Box 6887, Stanford, Calif. 94305)
>
> Seconded by: Henry M. Levin, regular member of *AEA* (address: School of Education, Stanford University, Stanford, California 94305)
>
> November 25, 1974

After discussion, the resolution PASSED by a vote of 37 to 27.

The second of the resolutions submitted by Behn and Levin read:

> We move that the American Economic Association adopt the following resolution:
>
> WHEREAS Federal laws regarding equal employment opportunity require open listing of job openings, there is no publication of the economics profession listing all job opportunities making such information easily accessible. While other professional associations maintain a listing for this purpose backed by a tradition of wide participation, this is not true for the American Economic Association. There is voluntary participation by members of the American Economic Association in the publication *Job Openings for Economists* (*JOE*), but there exists no tradition of open listing of job opportunities in one convenient location. The absence of open exchange of information on job openings is particularly true of the economics departments in the most prestigious institutions.
>
> WHEREAS the American Economic Association annual meeting schedule will be advanced in the next two calendar years to October, 1975 and September, 1976 it may be expected that the hiring process for economists will further deviate from competitive market conditions including the free exchange of information. At these early dates in the academic year, departments will not have formulated hiring plans in sufficient detail to permit identification of appropriate candidates so that greater reliance on the traditional faculty network to obtain names of potential employees may be expected, particularly for placement in academic positions. Therefore:
>
> BE IT RESOLVED that the American Economic Association alter the bimonthly publication schedule of *Job Openings for Economists* (*JOE*)—currently February, April, June, August, October and December—to add publication of *JOE* in September and November, 1975 and January and November in subsequent years. Further,
>
> BE IT RESOLVED that the recommendation of the American Economic Association Executive Committee (12–27–73) that "all members of the American Economic Association have a professional obligation to list their [job] openings with the Association" (i.e., in *JOE*); and the statement that "Employers are urged by the American Economic Association to allow sufficient time (30 days) for applicant response from the time of publication in *JOE* to the time the position is filled," shall be printed at the beginning of each issue of *JOE*.
>
> Proposed by: William H. Behn, regular

member of the *AEA* (address: P.O. Box 6887, Stanford, California 94305)

Seconded by: Henry M. Levin, regular member of the *AEA* (address: School of Education, Stanford University, Stanford, California 94305)

November 25, 1974

During the discussion, it was suggested that the first two paragraphs, each beginning "WHEREAS . . . ," be deleted. Behn and Levin agreed. The President ruled that the motion be divided with separate votes on each of the two remaining paragraphs. The Director of *Job Openings for Economists* objected to the first of the remaining paragraphs as premature. He said that experience was needed to ascertain whether more frequent publication of *JOE* would be worth the cost. Andrew Brimmer also spoke against this part of the resolution. Behn moved to table it. The motion to table was seconded and passed. After further discussion, in which the incoming President-elect Franco Modigliani spoke in favor of the remaining part of the resolution, it was PASSED. As passed it read:

> BE IT RESOLVED that the recommendation of the American Economic Association Executive Committee (12–27–73) that "all members of the American Economic Association have a professional obligation to list their [job] openings with the Association" (i.e., in *JOE*); and the statement that "Employers are urged by the American Economic Association to allow sufficient time (30 days) for applicant response from the time of publication in *JOE* to the time the position is filled" shall be printed at the beginning of each issue of *JOE*.

Behn and Levin's third resolution read:

We move that the American Economic Association adopt the following resolution:

WHEREAS Federal laws on equal employment opportunity provide no protection against political discrimination in employment, and whereas there is evidence of political discrimination in termination of employment of economists on university faculty, most recently at California State University, San Jose, and Lehman College (City University of New York), procedures should be developed by the American Economic Association to assist individual members of the Association who present allegations of political discrimination in employment. Therefore,

BE IT RESOLVED that the American Economic Association shall establish a standing committee, the Committee on Political Discrimination, whose members shall be appointed each year by the President of the Association to collect information on issues of fact in cases involving termination of employment for alleged political reasons. The Committee will be called into operation upon due application by the faculty member making the allegation of political discrimination. The committee will proceed with its investigation to the best of its ability, irrespective of the cooperation of the economics department in question, and within a reasonable time period make public its report. Due consideration shall be given by the Committee to the need for timely report so as to avoid extension of the case beyond the immediate academic year. A copy of the report shall be transmitted to the individual faculty member, his or her academic department, and the institution's president, as well as to the *AAUP* for consideration in censure proceedings against the university if the findings of fact indicate political discrimination. Furthermore,

BE IT RESOLVED that if the *AAUP* censures an institution, all job openings listed in *Job Openings for Economists* (*JOE*) for the six months prior to the censure shall be relisted in *JOE* with the censure duly noted. This censure notice shall appear in all subsequent listings from this institution in *JOE* until the censure is removed by the *AAUP*.

Proposed by: William H. Behn, regular member of the *AEA*. (address: P.O. Box 6887, Stanford, California 94305

Seconded by: Henry M. Levin, regular member of the *AEA*. (address: School of Education, Stanford University, Stanford, California 94305)

Behn asked permission to make two changes in the resolution: in the first of the paragraphs starting "BE IT RESOLVED . . . " to change "faculty member" to "academic economist," and to change "economics department" to "academic department." There was no objection. Several speakers discussed the situation at San Jose State, the response of the American Association of University Professors (*AAUP*) to it, and the situation at Lehman College. Robert Eisner moved to delete the words "and within a reasonable time make public its report." After the amendment had been seconded, Behn and Levin agreed to accept it. Responding to a suggestion by the President, Behn and Levin agreed to delete the words "most recently at California State University, San Jose, and Lehman College (City of New York)" from the paragraph beginning "WHEREAS . . . ." Fritz Machlup spoke against the Association making investigations, saying that, if it is not satisfied with the actions taken by the *AAUP*, its efforts should be directed toward getting the *AAUP* to do better. Bill Taub replied to Machlup's argument. Dorothy Crook suggested deleting the words "and whereas there is evidence of political discrimination on termination of employment of economists on university faculty" from the paragraph beginning "WHEREAS . . . ." Behn and Levin accepted the change. The Director of *JOE* pointed out that institutions censured by the *AAUP* are already listed in *JOE*. He moved that the entire last paragraph and the last word of the second last paragraph ("Furthermore") be deleted. The motion was seconded. Behn and Levin accepted the change. Modigliani raised questions about the resolution, particularly pointing out the problem of insuring due process. He suggested tabling the resolution to let the Executive Committee work out the problems. Brimmer said that the Executive Committee had discussed the resolution and was sympathetic to its spirit but was concerned about the difficulties with it. He said that as a result some members of the Executive Committee had drafted a revision of the resolution, which had been reproduced and distributed at the meeting with their names on it. He moved the revised version as a substitute motion. Machlup seconded the substitute motion, which read:

> WHEREAS the American Economic Association, at its Annual (Business) Meeting on December 29, 1972, adopted a resolution opposing political discrimination by economics departments in their hiring decisions, therefore:
>
> BE IT RESOLVED that the American Economic Association shall establish a standing committee, the Committee Against Political Discrimination. The Committee shall be receptive to any allegation of political discrimination brought to it by an academic economist. If the allegation appears to have some basis, the Committee shall then urge an investigation of the allegation by the *AAUP* or by any other appropriate body. The Committee shall make its concern known to the university and economics department in question, and it shall take a continuing, active interest in the case until its resolution.
>
> JOHN G. GURLEY
> PAUL M. SWEEZY

Brimmer proposed adding "or school" to the second last line after "economics department." Machlup consented. Machlup proposed striking the paragraph beginning "WHEREAS . . . " from the substitute motion on grounds that the paragraph concerned hiring whereas the problem concerned firing. Brimmer disagreed on grounds that "hiring practices" include firing practices. Brimmer and Machlup accepted a suggestion to add the phrase "and [new paragraph] WHEREAS Federal laws on equal employment opportunity provide no protection against political discrimination in employment" between "decisions" and "therefore" in the first paragraph of the substitute motion. Sweezy explained that he and Gurley had not moved the revised version of the resolution even though their names were on it because Behn and Levin did not agree to it. He said that he would abstain from voting on the motion to substitute the revised version and expected that Gurley would also abstain. Brimmer and Machlup accepted suggestions from Gary Fromm and the President to change "the university and economics department or school in question" to "appropriate officials and governing body of the organization in question" in the second paragraph of the substitute motion. Gurley proposed deleting the word "academic" from the second sentence of the second paragraph. Brimmer and Machlup at first agreed but then changed their minds. The previous question was moved, seconded, and passed. The motion to substitute the revised version as amended was then put to a vote and lost. The previous question was again moved, seconded, and passed. The original resolution as amended was then put to a vote and PASSED. As passed, it read:

> WHEREAS Federal laws on equal employment opportunity provide no protection against political discrimination in employment, procedures should be developed by the American Economic Association to assist individual members of the Association who present allegations of political discrimination in employment. Therefore,
>
> BE IT RESOLVED that the American Economic Association shall establish a standing committee, the Committee on Political Discrimination, whose members shall be appointed each year by the President of the Association to collect information on issues of fact in cases involving termination of employment for alleged political reasons. The Committee will be called into operation upon due application by the academic economist making the allegation of political discrimination. The committee will proceed with its investigation to the best of its ability, irrespective of the cooperation of the academic department in question. Due consideration shall be given by the Committee to the need for a timely report so as to avoid extension of the case beyond the immediate academic year. A copy of the report shall be transmitted to the individual faculty member, his or her academic department, and the institution's president, as well as to the *AAUP* for consideration in censure proceedings against the university if the findings of fact indicate political discrimination.

Having ascertained that Scheinman was now present, the President announced that he was adjourning the meeting until 9:30 p.m., when Scheinman's resolution would be considered. On resumption of the meeting, the President pointed out that the whole question of "unbundling," including permitting members to pay reduced dues and receive only one journal, was under consideration by the Executive Committee. Inasmuch as the marginal cost of each journal is very low relative to average cost, the reduction in dues would be small. Scheinman then read his resolution:

> Whereas, the American Economic Association, in its December 29, 1973, business meeting adopted a resolution considering 80 percent of the dues of annual members to be payment for a year's subscription to both the *American Economic Review* and the *Journal of Economic Literature;*
>
> Whereas, the American Economic Association does not currently have a nonpublications membership category;
>
> Whereas, the annual membership dues include payment for both publications and other activities of the Association;

Whereas, the number of annual members of the Association has declined 10 percent since its peak in 1970;

Whereas, annual dues in the Association have more than doubled since 1970;

Whereas, some members, whether for financial or other reasons, are unable to maintain their membership at the current annual rate and, therefore, must cancel their memberships either temporarily or permanently; and in so doing lose all benefits of affiliation with the Association;

Whereas, some members do not wish to receive the publications, but still wish to affiliate with the Association;

Therefore, be it resolved that the Executive Committee of the *AEA* shall initiate action to institute a new nonpublications membership category. A member in this category would not receive the *American Economic Review* or *Journal of Economic Literature*, but would receive all other membership benefits, including listing in and receipt of the membership directory, announcement of meetings, eligibility for voting in elections and eligibility for group insurance plans and charter flights. It is anticipated that this membership category carry a minimal charge of 20 percent of the annual membership dues.

Sincerely yours,
THEODORE P. SCHEINMAN

Seconded: H. Louis Stettler III

After some discussion, the resolution was put to a vote and lost.

There being no further business, the meeting was adjourned at 9:58 p.m.

RENDIGS FELS, *Secretary*

MINUTES OF THE EXECUTIVE COMMITTEE MEETINGS

Minutes of the Meeting of the Executive Committee in Washington, D.C., March 8, 1974.

The first meeting of the 1974 Executive Committee of the American Economic Association was called to order in the Washington Hilton Hotel, Washington, D.C., at 9:15 a.m. on March 8, 1974. The following members were present: Walter W. Heller, presiding, Kenneth J. Arrow, Gary S. Becker, George H. Borts, Andrew F. Brimmer, Rendigs Fels, R. A. Gordon, John G. Gurley, Robert L. Heilbroner, Guy Henderson Orcutt, Joseph A. Pechman, Mark Perlman, and Paul M. Sweezy. Irma Adelman and John Kenneth Galbraith were absent. Donald F. Turner was present as the Association's Counsel. Present as members of the Nominating Committee were James Tobin, Chairman, Allen C. Kelley, George L. Perry, Roy Radner, Christopher A. Sims, Alice Vandermeulen, and Charles Z. Wilson. Present as guests for parts of the meeting were Carolyn Shaw Bell, William C. Brainard, Robert Dorfman, Gary Fromm, Charles C. Holt, Barbara MacPhee, F. Ray Marshall, Barbara Reagan, and Robert C. Vowels.

1. *Minutes.* The minutes of the meeting of December 27, 1973, were approved as corrected.

2. *Report of the Secretary* (Fels). The Secretary reported on future meeting plans. It was VOTED to change the 1976 meetings from September 16–18 in Atlantic City to December 28–30 in Philadelphia.[1] It was agreed that the preferred dates for the meetings scheduled for 1980 in Denver were September 5–7 (Friday through Sunday), with the meeting of the Executive Committee on September 4. The Secretary reported that the questionnaires for the 1974 *Directory of Members* (formerly called the *Handbook*) had been mailed to all members. Production of the *Directory* will be computerized. The computer tape will subsequently be available for preparation of special listings. It was VOTED to establish an *ad hoc* committee on computerization of the Association's mailing list and payment records. On recommendation of the Secretary it was VOTED to cosponsor the research project of the American Political Science Association on confidentiality of research sources. It was agreed that the post of Assistant Secretary-Treasurer would henceforth be halftime during the academic year.

3. *Report of the Treasurer* (Fels). The Treasurer reported that preliminary figures from the auditors for 1973 indicated a deficit of $15,000. The deficit resulted in part from real capital losses of $127,000, $39,000 of which came from the rise in the *GNP* deflator with most of the remainder from the fall in the stock market. It was VOTED to increase dues of annual members $2 (to $23) and other dues and subscriptions proportionately effective January 1, 1975. It was VOTED to require a report from the Budget Committee in the future before considering dues increases. It was VOTED to ask the Budget Committee to study the possibility of greater progressivity in the dues structure and to report its findings to the Executive Committee on December 27, 1974. In accordance with reports described in paragraphs 4, 10, and 11 below, it was VOTED to increase the budget for 1974 by a total of $9,000, including $3,000 for the *American Economic*

---

[1] This decision was subsequently reversed by mail ballot.

Review, $5,000 for startup costs for *Job Openings for Economists*, and $1,000 for a questionnaire. It was VOTED to repeal the directive adopted by the Executive Committee at the meeting of March 11–12, 1966, establishing the policy of investing in equities all funds not needed for current operating expenses or for reserves for special grants ("Minutes of the Executive Committee Meetings") and to ask the Finance Committee to report to the next meeting of the Executive Committee its recommendation on what the policy should be.

4. *Report of the Editor of the American Economic Review* (Borts). The Managing Editor reported on his efforts to add to the Board of Editors a radical economist. It was VOTED to authorize him to appoint any one of four individuals proposed by the Executive Committee. In accordance with the Managing Editor's recommendation, it was VOTED to increase the 1974 budget for the *Review* by $3,000 to provide temporary secretarial assistance.

5. *Report of the Editor of the Journal of Economic Literature* (Perlman). The Managing Editor reported on problems relating to computerization, proliferation of economics journals and its effect on the *Journal's* costs, and accounting.

6. *Report of the Committee on the Journals* (Brainard). The Chairman reported on ways that the Association might support *Challenge* magazine. After considerable discussion, it was agreed that, while the Executive Committee believes *Challenge* meets an important need and *AEA* members have a stake in its success, any action taken should be by members of the Executive Committee in their individual capacities, since *Challenge* is a commercial venture with good prospects of success.

The Chairman recommended creation of a small committee to continuously evaluate the journals. Its function would differ from that of the boards of editors, which assist the managing editors with evaluation of manuscripts, and from that of the Budget Committee, which reviews the budgeting recommendations of the managing editors. The new committee would be concerned with overall policy and with how well the Association was meeting the needs of its members. If the managing editors need assistance on day-to-day operations, they can employ consultants. It was VOTED to establish a standing Committee on Publications.

The Chairman proposed that a questionnaire be sent to a 10 percent sample of the membership to find out how members use the journals, what their attitude is toward questions of future policy, and how well served they feel as authors or potential authors. It was agreed that this proposal should be considered a suggestion to the Committee on Publications, which was authorized to proceed with the questionnaire at its discretion.

7. *Report of the Counsel* (Turner). The Counsel discussed various ways of improving the handling of resolutions offered at the business meeting. It was VOTED to require that proposed resolutions be submitted to the Secretary at least one month in advance of the annual meeting in writing with signatures of the proposer and a second, both of whom must be members in good standing of the Association, such resolutions to be made available ahead of time. It was agreed that notice of this rule would be included with the mailing of ballots in the fall of 1974. A suggestion was made that the President announce at the beginning of the Annual Meeting his rulings on whether proposed resolutions were in order, that any appeal to his rulings be debated and voted on at that time, but that consideration of the resolutions themselves be at the end of the meeting. It was agreed that the decision on this question of timing would be left to the Counsel, the President, the President-elect,

and the Secretary. To be able to adjourn the Annual Meeting at a reasonable hour, it was agreed to schedule the Presidential Address at 7:30 p.m. instead of 8:00 p.m.

The Counsel advised clarification of the bylaw requiring that members be sent all publications of the Association. In his opinion, the bylaw refers to the *American Economic Review*, the *Journal of Economic Literature*, and the *Directory of Members* only; but in view of the proposal (see paragraph 10 below) for a new publication to be called *Job Openings for Economists*, he proposed that Paragraph 4 of Section I of the bylaws be amended to read as follows:

> 4. Every member is entitled to receive, as they appear, the following publications of the Association: the *American Economic Review*, the *Journal of Economic Literature*, and the *Directory of Members*.

It was VOTED to approve this amendment and to instruct the Secretary to submit it to the members for approval by mail ballot.[2]

8. *Nominating Committee* (Tobin). The Chairman reported the following nominees for offices in the 1974 election: for Vice-President (two to be chosen), Martin Bronfenbrenner, James Morgan, Charles Schultze, and Leonid Hurwicz; members of the Executive Committee (two to be chosen), Jack Hirschleifer, Carolyn Shaw Bell, Burton A. Weisbrod, and Dale Jorgenson. The Electoral College, consisting of the Nominating Committee and the Executive Committee meeting together, chose William Parker as representative to the American Council of Learned Societies, Orcutt as representative to the Social Science Research Council, and Franco Modigliani as nominee for President-elect.

9. *Program Committee* (Gordon). The President-elect reported on his plans for the program of the 1974 meetings. He also reported that he had received a letter from Thomas Smith concerning the bicentennial of Adam Smith and suggested that there be a session on the latter at the meetings in 1975 or 1976.

10. *Committee on the Status of Women in the Economics Profession* (Bell). It was VOTED that the Committee on the Status of Women in the Economics Profession henceforth be a standing committee of the Association. It was further VOTED to authorize the Committee to accept and expend funds for specified projects falling within its mandate with the consent of the President.

11. *Committee on Hiring Practices* (Marshall). It was VOTED to adopt the following recommendations submitted by the Chairman to implement the decision made at the previous meeting to publish job lists for economists:

> (1) The job list, to be called *Job Openings for Economists* (*JOE*), should be issued six times a year, in accordance with the committee's recommendation, as approved by the Executive Committee at its December 27, 1973, meeting.
>
> (2) The annual charge for *Job Openings for Economists* should be set at $6 per year. Subscriptions should be made available to individual members and to departments and to other organizations.
>
> (3) The President of the *AEA* should send a letter to all members of the association in April 1974 informing members about *Job Openings for Economists*, soliciting subscriptions, and alerting chairmen that they will be solicited for job listings.
>
> (4) The President should follow his letter to the membership with one to economics department chairmen, asking them to list their openings in *Job Openings for Economists*. This request should go out no later than August 15, setting September 15 as the deadline for information. This letter should contain a statement of the association's open listing policy and urge department chairmen in the strongest possible terms to

---

[2] The amendment was ratified by the members in June 1974.

list their openings. The cooperation of departments is absolutely essential to the success of this effort.

(5) An effort should be made to get the first copy of *Job Openings for Economists* in the mail by October 1, 1974, and at least by December 1, 1974.

(6) The publication of *Job Openings for Economists* should be governed by the following policies:

(a) Only members of the *AEA* may subscribe to *JOE*.
(b) No listing will be accepted containing material which violates federal civil rights laws.
(c) Listings will be accepted from institutions under censure by the *AAUP*, but the censured institutions will be noted.
(d) Jobs will be listed according to the classification of fields currently used by the *Journal of Economic Literature*.

12. *Job Market Questionnaire*. Bell and Marshall reported that a joint questionnaire to be sent to department chairmen had been drafted to provide information on the job market for the Committee on the Status of Women in the Economics Profession, *Job Openings for Economists*, the Committee on Minority Group Economists, and department chairmen. The draft needed further revision and editing. It was VOTED to approve the plan to collect data through such a questionnaire.

13. *Committee on Minority Group Economists* (Dorfman). The Chairman reported that a Summer Program for Minority Group Economists under the direction of Marcus Alexis would be held in 1974 on the campus of the University of California at Berkeley. Robert Vowels voiced the support of the Caucus of Black Economists for the summer program. The Executive Committee gave firm endorsement to the program, and various members present offered to help promote applications from students. It was VOTED to change the special Committee on Minority Group Economists to a standing Committee on the Status of Minority Groups in the Economics Profession to work with the Caucus of Black Economists and any similar organizations representing minority groups.

14. *United States-USSR Economic Conference* (Heller). The President reported on plans for colloquia of *U.S.* and *USSR* economists. An application for a $7,000 grant to the *AEA* to host the first meeting in the United States next fall or winter has been made to *IREX*. When the grant is received, the President will appoint a small committee to administer the first conference.

15. *Proposal for Accreditation*. A letter dated February 14, 1974, from Ali M. S. Fatemi to the President had been circulated to the Executive Committee in accordance with the discussion ("Minutes of the Annual Meeting, December 29, 1973") at the Annual Meeting. The letter proposed that a committee be appointed to consider taking an active role in accrediting economics departments or to establish professional standards. No action was taken.

16. *Census Advisory Committee* (Fromm). Fromm submitted for consideration three proposed resolutions on confidentiality of census data. No action was taken.

17. *Policy-Oriented Research* (Holt). Holt submitted for consideration a proposed resolution recommending that the federal government direct policy-oriented research toward national problems. It was agreed to put the resolution on the agenda of the meeting scheduled for December 27, 1974.

18. *Date of Spring Meeting*. It was agreed that the Executive Committee would meet on March 14–15, 1975, in Washington, D.C.

The meeting adjourned at 5:35 p.m.

Minutes of the Meeting of the Executive Committee in San Francisco, California, December 27, 1974.

The second meeting of the 1974 Executive Committee of the American Economic Association was called to order in the San Francisco Hilton Hotel, at 10:15 a.m. on December 27, 1974. The following members were present: Walter W. Heller, presiding, Irma G. Adelman, George H. Borts, Andrew F. Brimmer, Rendigs Fels, John Kenneth Galbraith, Robert Aaron Gordon, John G. Gurley, Robert L. Heilbroner, Joseph A. Pechman, Mark Perlman, and Paul M. Sweezy. Also present were newly elected members Carolyn Shaw Bell, Martin Bronfenbrenner, Franco Modigliani, Charles L. Schultze, and Burton A. Weisbrod. Present as Counsel was Donald F. Turner. Absent were Kenneth J. Arrow, Gary S. Becker, and Guy Henderson Orcutt. Present as guests for parts of the meeting were William C. Brainard, C. Elton Hinshaw, Fritz Machlup, Barbara MacPhee, and Barbara B. Reagan.

1. *Minutes*. The minutes of the meeting of March 8, 1974, were approved as corrected.

2. *Report of the Secretary* (Fels). The Secretary reported on future meeting plans. It was VOTED to meet on the east coast in September 1982 and in San Francisco during the Christmas holidays in 1983. For 1982, the Secretary was asked to investigate New York, Boston, and Montreal. It was VOTED to approve a contract dated February 4, 1974, with Hilton Hotels Corporation covering 1977–81 except for the paragraph on pages 3–4 concerning the 1975 meetings in Dallas.

3. *Report of the Treasurer* (Fels). The Treasurer presented a report including a budget for 1975 incorporating the recommendations of the Budget Committee. Pechman criticized the form of the budget, saying there should be three statements (cash flow, accrual, and assets and liabilities) instead of one statement mixing three kinds of information. Modigliani asked for discussion of the price of the *Index*. It was the sense of the meeting that a prepublication price of $10 be offered to members with the price otherwise set at $15 (or higher) and that the Secretary-Treasurer should explore with Richard D. Irwin, Inc., distributor of the *Index*, the suggestion to promote sales of the *Index* to economics departments and possibly to offer them a discount for purchasing all four of the volumes planned for publication in 1975. After some discussion, it was agreed to postpone a decision on when to publish a new *Directory*, in the meantime continuing to keep $50,000 a year (or a similar amount) in the budget as a reserve for it. The Chairperson of the Committee on the Status of Women in the Economics Profession (Reagan) asked that the Executive Committee appropriate $12,700 for the Committee for 1975 instead of the $5,000 in the budget. It was VOTED to appropriate $10,000. It was VOTED to raise the submission fee for the *American Economic Review* from $10 to $15. It was VOTED to raise the subscription price for *Job Openings for Economists* (*JOE*) to $12 per year except for junior members, for whom it will remain at $6. There was discussion of the policy of not paying the expenses of elected members of the Executive Committee to the annual meetings, a policy based on the expectation that their institutions will pay their expenses. It was VOTED to pay the expenses of a current member of the Executive Committee who is self-employed with the understanding that this action would not be regarded as a precedent for making other kinds of exceptions. It was then VOTED to approve the budget as amended. It was VOTED to raise dues effective January 1, 1976, to $25 for regular members, with proportional increases for other membership classes, subject to reconsideration at later meetings of the Executive Committee. It was agreed to defer until the March 1975 meeting the question of increasing the progressivity of the dues structure. In accordance with the sense of the meeting, Fels and Brimmer determined the following wording for a directive to the Finance

Committee to replace the one repealed last March:

> Funds of the Association not needed for current operating expenses or reserves for special funds shall be invested with due regard for yield, liquidity, and safety.

4. *Report of the Editor of the American Economic Review* (Borts). On recommendation of the Managing Editor of the *American Economic Review*, it was VOTED to renominate Anna Schwartz to the Board of Editors and to nominate Eugene Fama, Robert J. Gordon, James Melvin, and William Nordhaus. The Editor reported that 25 percent of the authors of articles in the December 1974 issue of the *Review* were not members of the Association. He inquired about the appropriateness of his writing to them suggesting that they become members. It was agreed that such letters should be sent by the Secretary.

5. *Report of the Editor of the Journal of Economic Literature* (Perlman). On recommendation of the Managing Editor of the *Journal of Economic Literature*, it was voted to nominate Moses Abramovitz, Arthur Goldberger, Thomas Mayer, and Ryuzo Sato to the Board of Editors.

6. *Report of the Committee on Publications* (Brainard). The Chairperson reported that the Committee on Publications had decided to proceed with the market survey authorized last spring. Although, as part of a research project for the National Science Foundation, Machlup will make a pilot study of the extent to which journal articles are read, which will be useful to the Committee, the study will not yield all the specific information needed by the Association to make decisions about cutting costs. Machlup requested free use of the Association's mailing list for his study. It was agreed to grant the request. The President-elect asked that the two editors report to the March meeting what they would do if required to cut costs 10 percent. Modigliani suggested polling the members on whether they would be willing to pay more to have the *Directory* published every three years instead of every four or five. Brainard said that the survey to be carried out by the Committee on Publications could include a question on this matter.

7. *Report of the Director of Job Openings for Economists* (Hinshaw). The Director of *Job Openings for Economists* (*JOE*) reported that the new publication was off to a good start. Bell asked that the Executive Committee be given a separate accounting of the costs and revenues of *JOE*.

8. *Report of the Chairman of the 1975 Program Committee* (Modigliani). The Chairman of the 1975 Program Committee reported on his plans. There was discussion of the advisability of the October dates for the 1975 meetings. It was the sense of the meeting that the dates should be shifted to the traditional period between Christmas and New Year's if feasible. The President appointed a committee to investigate the feasibility of such a shift consisting of Fels as chairman, Galbraith, Ralph Green, MacPhee, Modigliani, and Reagan.

9. *Report of the Counsel* (Turner). The Counsel gave his opinion as to which of the resolutions to be considered at the Annual Meeting on December 29, 1974, were in order. He also stated that an amendment within the subject matter of the resolution would be in order. It was VOTED to extend his term of office.

The meeting was adjourned at 5:10 p.m.

RENDIGS FELS, *Secretary*

REFERENCES

"Minutes of the Executive Committee Meetings," *Amer. Econ. Rev. Proc.*, May 1967, *57*, 681.

"Minutes of the Annual Meeting, New York, N.Y., December 29, 1973," *Amer. Econ. Rev. Proc.*, May 1974, *64*, 446–47.

REPORT OF THE SECRETARY FOR 1974

Annual Meetings. At the request of the Executive Committee, the dates of the 1975 annual meetings in Dallas have been changed from early October to December 28–30. The Employment Center will begin operations December 27 to give those concerned a greater opportunity to attend the sessions scheduled for December 28–30. The schedule for subsequent meetings is: September 16–18, 1976, Atlantic City, Chalfonte-Haddon Hall; December 28–30, 1977, New York, New York Hilton Hotel; August 29–31, 1978, Chicago, Conrad Hilton Hotel; December 28–30, 1979, Atlanta, Atlanta Hilton and Peachtree Center Plaza; September 5–7, 1980, Denver, Denver Hilton Hotel; and December 28–30, 1981, Washington, D.C., Washington Hilton Hotel and Sheraton-Park Hotel. The 1982 meetings will be held in the vicinity of Labor Day at an eastern city to be determined. Negotiations are underway with a view to meeting in San Francisco, December 28–30, 1983.

Membership. Table 1 shows that at the end of 1974 the total number of members and subscribers had nearly regained the peak reached at the end of 1970 just before dues and subscriptions were doubled in price. Even though the price of subscriptions was subsequently raised another 55 percent, the number of subscribers is now greater than in 1970. In contrast, though the subsequent increase in dues was only 5 percent, the number of members is still below the peak. Demand is inelastic for both members and subscribers, especially subscribers. The fall has been especially marked for annual members. The number of junior members, whose dues are only half as great as the dues of annual members, has increased 50 percent.

TABLE 1—MEMBERS AND SUBSCRIBERS
(End of Year)

	1970	1973	1974
Class of Membership			
Annual	16,075	14,453	14,544
Junior	2,077	2,480	3,135
Family	289	277	313
Complimentary	160	413	418
Life	291	332	342
Honorary	16	14	14
Total Members	18,908	17,969	18,766
Subscribers	7,265	7,228	7,353
Total Members and Subscribers	26,173	25,197	26,119

Directory. The 1974 *Directory of Members*, formerly called the *Handbook*, was published and distributed to members late in the year. It contains biographical and bibliographical information. The information is on a computer tape, and a program has been written to make possible the preparation of special lists for special purposes.

Employment Services. The National Registry for Economists continues to be operated on a year-round basis by the Illinois State Employment Service under the direction of Mrs. Theresa Scholl. This is an automated retrieval service in which the anonymity of employers is maintained. All economists looking for jobs are urged to register. The Association is indebted to Mrs. Scholl not only for the Registry but also for her expert assistance and supervision of the employment service provided at the annual meetings of the Allied Social Science Associations by the state offices of the U.S. Employment Service. During 1974, the Association inaugurated publication of *Job Openings for Economists* and discontinued the "Vacancies and Applications" section of the *American Economic Review.*

Permission to Reprint and Translate. Official permissions to quote from, reprint, or translate and reprint articles from the *American Economic Review* and the *Journal of Economic Literature* totaled 180 in 1974 compared to 147 in 1973. In addition permission was granted to reprint 22 abstracts from the *Journal.* Upon receipt of a request for permission to reprint an article, the publisher or editor making the request is instructed to get the author's permission in writing and send a copy to the Secretary as a condition for official permission. The Association suggests that authors charge a fee of $150, but they may charge some other amount, enter into a royalty arrangement, waive the fee, or refuse permission altogether.

Visiting Economics Scholars Program. A Visiting Economics Scholars Program, modeled on the Visiting Scientists Program in Economics formerly supported by the National Science Foundation, was inaugurated for the academic year 1974–75 under the direction of the Assistant Secretary, C. Elton Hinshaw. Its purpose is to facilitate visits by leading economists to smaller colleges emphasizing teaching. The colleges are expected to pay part or all of the costs of the visits in accordance with their financial resources; at a minimum they take care of the local expenses of the visitor. Inquiries should be directed to the Assistant Secretary.

Committees and Representatives. Listed below are those who served the Association during 1974 as members of appointed committees or as representatives. Years in parentheses indicate the final year of the term to which they most recently have been appointed. On behalf of the Association, I wish to thank all of them for their service.

Standing Committees

ADVISORY COMMITTEE ON THE HISTORY OF THE ASSOCIATION

George J. Stigler, *Chairperson*
Joseph Dorfman
Harold F. Williamson, *Corresponding Secretary*

BUDGET COMMITTEE

Joseph A. Pechman, *Chairperson* (1974)
Irma Adelman, *Chairperson* (1975)
Andrew F. Brimmer (1976)
*Burton A. Weisbrod (1977)
Rendigs Fels, *Ex Officio*
Walter W. Heller, *Ex Officio* (1974)
R. A. Gordon, *Ex Officio* (1975)
*Franco Modigliani, *Ex Officio* (1976)

ADVISORY COMMITTEE ON STUDIES OF THE LABOR MARKET FOR ECONOMISTS

F. Ray Marshall, *Chairperson*
Barbara Reagan
T. Aldrich Finegan
Francis M. Boddy

CENSUS ADVISORY COMMITTEE

Lee E. Preston (1977), *Chairperson* (1974)
Richard Ruggles, *Chairperson* (1975)
Guy E. Noyes (1974)
Patricia Shontz (1974)
Joseph Burns (1975)
Anne P. Carter (1975)
Dean E. McKee (1975)

---

* Newly appointed in 1975.

J. Fred Weston (1975)
Gardner Ackley (1976)
Armen Alchian (1976)
Andrew F. Brimmer (1976)
Anthony Downs (1976)
James R. Nelson (1976)
Jacob Mincer (1977)
*George L. Perry (1977)
Phyllis Wallace (1977)

COMMITTEE ON ECONOMIC EDUCATION

G. L. Bach, *Chairperson* (1976)
Elisabeth Allison (1975)
Kenneth E. Boulding (1975)
Allen C. Kelley (1975)
Henry H. Villard (1976)
Phillip Saunders (1977)
Rendigs Fels, *Ex Officio*

COMMITTEE ON HONORARY MEMBERS

Fritz Machlup, *Chairperson* (1976)
Lawrence R. Klein (1976)
Bent Hansen (1978)
W. Arthur Lewis (1978)
Leonid Hurwicz (1980)
Paul A. Samuelson (1980)

COMMITTEE ON HONORS AND AWARDS

Robert M. Solow, *Chairperson* (1974)
Lloyd G. Reynolds, *Chairperson* (1976)
Gary Becker (1976)
Irma Adelman (1978)
Marcus Alexis (1978)
*John Chipman (1980)
*James W. McKie (1980)

COMMITTEE ON THE STATUS OF MINORITY GROUPS IN THE ECONOMICS PROFESSION

Marcus Alexis, *Chairperson* (1977)
Bernard E. Anderson (1975)
Karl Gregory (1975)
Phyllis Wallace (1975)
George Borts (1976)
Andrew Brimmer (1976)
Alice Rivlin (1976)
James Tobin (1977)
Charles Z. Wilson (1977)

COMMITTEE ON POLITICAL DISCRIMINATION

Kenneth J. Arrow, *Chairperson*
William J. Baumol
John G. Gurley
Anne Krueger
F. Ray Marshall
Carl Stevens
Thomas E. Weisskopf

COMMITTEE ON PUBLICATIONS

William Brainard, *Chairperson* (1977)
John G. Gurley (1975)
Robert Heilbroner (1975)
Peter Diamond (1976)
Robert Lampman (1976)
Michael Lovell (1977)

COMMITTEE ON THE STATUS OF WOMEN IN THE ECONOMICS PROFESSION

Carolyn Shaw Bell (1975), *Chairperson* (1974)
Barbara Reagan, *Chairperson* (1975)
Walter Adams (1975)
Francine Blau (1975)
Martha Blaxall (1975)
Kenneth Boulding (1975)
Collette Moser (1975)
Myra Strober (1975)
Phyllis Wallace (1975)
Florence Weiss (1975)
Walter W. Heller, *Ex Officio* (1974)

FINANCE COMMITTEE

Beryl W. Sprinkel, *Chairperson*
C. Wells Farnham
Rendigs Fels
Milton Friedman
Walter W. Heller

---

* Newly appointed in 1975.

ECONOMICS INSTITUTE POLICY AND ADVISORY BOARD

Edwin S. Mills, *Chairperson* (1975)
Henry Rosovsky (1974)
James Worley (1974)
Irma G. Adelman (1975)
Walter P. Falcon (1976)
Daniel Schydlowsky (1976)
Arnold Harberger (1977)
Richard H. Holton (1977)

JOINT COMMITTEE WITH THE ASSOCIATION OF AMERICAN LAW SCHOOLS

George J. Stigler, *Co-Chairman* (1976)
Louis De Alessi (1974)
Harold Demsetz (1974)
Gerald M. Meier (1975)
Peter O. Steiner (1975)
Alvin Klevorick (1976)

Special Committees

COMMITTEE ON COMPUTERIZATION

John R. Meyer, *Chairperson*

COMMITTEE ON HIRING PRACTICES

F. Ray Marshall, *Chairperson*
Carolyn Shaw Bell
Robert J. Brown
Allan M. Cartter
Stephen Marglin

NOMINATING COMMITTEE (1974)

James Tobin, *Chairperson*
Allen C. Kelley
George L. Perry
Roy Radner
Christopher A. Sims
Alice Vandermeulen
Charles Z. Wilson

NOMINATING COMMITTEE (1975)

Robert Eisner, *Chairperson*
Bernard E. Anderson
Robert E. Gallman
Richard W. Parks
L. A. Rapping
Barbara B. Reagan
Christopher A. Sims

COMMITTEE ON ELECTIONS (1974)

Wilfred L. David, *Chairperson*
Malcolm Getz
Rendigs Fels, *Ex Officio*

ADVISORY COMMITTEE ON MINORITY GROUP ECONOMISTS

Robert Dorfman, *Chairperson*
Marcus Alexis
George H. Borts
Andrew Brimmer
Karl Gregory
Charles Schultze
Thaddeus Spratlen
James Tobin

COMMITTEE ON JOURNALS

William Brainard, *Chairperson*
Mary Jean Bowman
Robert Fogel
John G. Gurley
Burton A. Weisbrod

SAN FRANCISCO LOCAL ARRANGEMENTS COMMITTEE (1974)

Richard H. Holton, *Chairperson*
Leslie Carbert
Kent Sims
Barbara MacPhee
W. D. Hermann
F. Theodore Malm
Theresa Scholl

Charles Roumasset
Edward Smith
Jerry Faulkner
Ted Gibson
Sandra Peek
Constance Holton
Gene Conatser
Noralyn Ewart
Thomas H. Carmody
Dennis MacDonneil
Charlotte Alhadeff
Jeanette Sims
Harold L. Buma
R. Alan Carl
Diana Blank
Lee J. Nash
John O. Youngs

DALLAS LOCAL ARRANGEMENTS COMMITTEE (1975)

Ralph T. Green, *Chairperson*
Robert Smith, III
J. Carter Murphy
Paul Weathers
Barbara A. MacPhee
Jack Strickland
Douglas Foster
Eugene C. Zorn, Jr.
George W. Kelly
Josef Hadar
James A. Byrd
T. E. McMillan, Jr.
S. T. Keim, Jr.
Hilary S. Ball
Thomas P. Robertson

Council and Other Representatives

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

Stephen Goldfeld (1976)

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SLAVIC STUDIES

Janet Chapman (1976)

AMERICAN COUNCIL OF LEARNED SOCIETIES

William Parker (1978)

AMERICAN POLITICAL SCIENCE ASSOCIATION—JOINT RESEARCH PROJECT ON CONFIDENTIALITY OF RESEARCH SOURCES

Gary Fromm

FEDERAL STATISTICS USERS' CONFERENCE

John W. Kendrick (1977)

INTERNATIONAL ECONOMIC ASSOCIATION

Fritz Machlup (1975)
Abram Bergson (1978)

INTERSOCIETY COMMITTEE ON TRANSPORTATION

Charles A. Taff (1977)

JOURNAL OF RESEARCH ON CONSUMER BEHAVIOR

Kelvin J. Lancaster (1975)

NATIONAL ARCHIVES ADVISORY COUNCIL—GENERAL SERVICES ADMINISTRATION

Robert Gallman (1975)

NATIONAL BUREAU OF ECONOMIC RESEARCH

Willard L. Thorp (1975)

NATIONAL COUNCIL FOR ACCREDITATION OF TEACHER EDUCATION

Phillip Saunders (1977)

NATIONAL RESEARCH COUNCIL—DIVISION OF BEHAVIORAL SCIENCES

Harold W. Watts (1974)
Barbara Bergmann (1975)

SOCIAL SCIENCE RESEARCH COUNCIL

James N. Morgan (1974)
William J. Baumol (1975)
Lawrence R. Klein (1976)
*Guy Orcutt (1977)

Representatives of the Association on Various Occasions—1974

INAUGURATIONS

Lee H. Smith, Southwest Texas State University
Forest H. Hill

William Wayne Jellema, Wartburg College
B. Wylie Anderson

Harold M. Proshansky, The Graduate School and University Center, City University of New York
Henry H. Villard

Robert Paul Lisensky, Willamette University
Richard A. Munn

RENDIGS FELS, *Secretary*

---

* Newly appointed in 1975.

REPORT OF THE TREASURER FOR THE YEAR ENDING DECEMBER 31, 1974

The table on the next page gives budget data for 1973, 1974, and 1975. The budgets are compared with the actual results for 1973 and with the results anticipated for 1974 at the time (December 27, 1974) that the 1975 budget was approved by the Executive Committee. The actual results for 1974 are given below in the financial statements following the auditor's report. Since the Treasurer's Report is sent to the printer before the audit takes place, the final results cannot be known at the time this report is written.[1] Ordinarily actual results differ from anticipated by a significant margin. Nevertheless, it is clear that a very large deficit was incurred in 1974 and an even larger one is in prospect for 1975.

The 1974 deficit resulted from a shortfall of investment income from the decline in the stock market, spending on the periodicals above the budgeted amounts because of unexpectedly large increases in paper and printing costs, the decisions by the Executive Committee in March 1974 to launch a new publication (*Job Openings for Economists*) and make the post of Assistant Secretary-Treasurer half-time, and a planned deficit of $10,000 in the budget adopted by the Executive Committee on December 27, 1973.

The most important reason for the deficit was real capital losses. Under the formula used by the Association for calculating income from investments, real capital gains and losses are included whether realized or not. To the extent that gains or losses are from stocks, they are spread over three years. The net loss for 1974 includes a gain of $25,069 carried forward from 1972; a loss of $58,509 carried forward from 1973; part of the real capital losses incurred during 1974 (the part to be recognized in 1974); and dividends and interest. The rise in paper and printing costs is second only to the investment losses in accounting for the deficit. Some increase had been anticipated in the budget approved on December 27, 1973, but the actual increase turned out to be much greater.

The line item for *Job Openings for Economists* (*JOE*) understates its cost because it does not include salaries. *JOE* is eventually expected to become self-supporting, but start-up costs and temporary subsidization of operating expenses contributed about $12,000 to the expected deficit for 1974.

The projections of revenues and the proposals for expenditures for 1975 indicate a deficit on the order of $155,000.[2] There is no provision in the budget for contingencies. Actual results are usually worse than expected.

The expected deficit of $155,000 is on what might be called an accrual basis. The figure includes $99,000 of investment losses to be carried forward from 1973 and 1974. It also includes a reserve for a new *Directory* in 1977 (or some later year) of $50,000. On the other hand, it omits $55,000 representing the manufacturing cost of four volumes of the *Index of Economic Articles* scheduled for publication in 1975. It is expected that the manufacturing costs, which will be borne by the Association, will be recovered in subsequent years by sales. It therefore has not been included in the budget shown in Table 1, but it will require a cash outlay.

---

[1] Some changes, however, have been made in galley proof, including the addition of this and other footnotes and revision of the projections of revenues in Table 1.

[2] At its meeting on March 14, 1975, the Executive Committee took a number of actions to improve the financial position of the Association. Although many of them will have no effect until 1976, the size of the expected deficit for 1975 would be markedly smaller if these actions were taken into account.

TABLE 1—1975 BUDGET OF THE AMERICAN ECONOMIC ASSOCIATION
Compared to 1973 and 1974
(thousands of dollars)

	1973		1974		1975
	Budget	Actual	Budget[a]	Projection[b]	Budget[c]
REVENUES					
Dues and subscriptions	500	506	530	583	620
Advertising	60	73	80	67	61
Sales—miscellaneous	20	18	15	21	15
Sales—mailing list	14	20	25	23	24
Annual meeting	13	7	15	9	10
Job Openings for Economists	—	—	—	4	15
Other	13	11	14	13	18
Subtotal	620	636	679	720	763
INVESTMENT GAINS (LOSSES)	19	1	19	(55)	(56)
TOTAL REVENUES	639	637	698	665	707
EXPENSES					
Publications					
American Economic Review	165	172	178	198	232
Journal of Economic Literature and Index	220	221	235	249	274
Papers and Proceedings	47	45	48	49	58
Job Openings for Economists	—	—	—	10	15
Directory	40	40	70	65	50
Subtotal	472	479	531	571	629
General and administrative					
Committees	16	17	15	15	25
Salaries	88	86	96	103	123
Rent	8	8	8	8	8
Other	60	63	58	67	77
Subtotal	172	174	177	193	233
TOTAL EXPENSES	644	653	708	764	862
REVENUES LESS EXPENSES	(5)	(15)	(10)	(99)	(155)

[a] As approved by the Executive Committee, December 17, 1973. Does not include additional expenditures authorized March 14, 1974.

[b] Estimates prepared November 5, 1974, of revenues and expenses for the entire year.

[c] The projections of revenues for 1975 were revised in March in the light of the actual results for 1974. The estimates of expenditures take account of the actions by the Executive Committee at its meeting of December 27, 1974, and by the membership at the Annual Meeting of December 29, 1974. Actions taken by the Executive Committee at its meeting on March 14, 1975, however, have not been incorporated in the estimates of revenues and expenditures.

Adjusting for these three items (investment losses of past years, the reserve for the *Directory*, and the manufacturing cost of the *Index*) results in a cash flow deficit of $76,000.

Neither of the two deficits estimated above measures the change in the net worth of the Association. Since the Association is not under obligation to publish a new *Directory* in 1977 or any other year, setting aside a reserve of $50,000 does not change the net worth. Neither does the $55,000 to be spent on manufacturing the *Index* (provided the money is recovered with interest from sales of future years). The investment losses carried forward from 1973 and 1974 likewise do not represent a change in the net worth of the Association between January 1 and December 31, 1975, even though they are recognized in that year. Therefore, the net worth of the Association is expected to rise (not fall) in 1975.

The accrual deficit, the cash deficit, and the rise in net worth are all relevant to the financial problems of the Association but in different ways. Of the three, the accrual deficit is the one of greatest importance for budget decisions. The practice of recognizing only one-third of the real capital losses incurred from equities in the year in which they occur allows for the vagaries of the stock market. The losses of 1973 (and 1974) not recognized in that year did not affect that year's budget decisions, but the corollary is that they should affect later decisions. Although there is no legal obligation to publish another *Directory*, spending decisions must take account of the likelihood that some future Executive Committee will want one.

The size of the deficits that may be regarded with equanimity depends in part on the ratio of net worth to annual expenditures. At the end of 1973, the net worth was $162,338, not counting $80,000 in the reserve for the 1974 *Directory* which has now been spent (but was not counted as part of the 1974 deficit). The ratio of this figure to the expenditures of 1974 is less than 0.3. Since then the net worth has gone down. It was negative at the end of 1974.

Another criterion for the financial position is the size of the net worth relative to the largest previous deficit. By this test, the deficits of 1974–75 are excessive. The largest deficit ever incurred by the Association was in 1970. It was $242,000—a figure in excess of the net worth at the end of 1973.

In my opinion, the Association should aim for surpluses on the order of $100,000 in 1976 and later years until its net worth rises to an adequate ratio to expenditures. Continuing inflation will raise annual expenditures year by year even if no new programs are undertaken and no old ones expanded. With an annual budget in excess of $800,000 a year and an inflation rate of 5 percent, the net worth would need to rise $20,000 a year even if it were initially adequate, which it is not; and with an inflation rate of 10 percent, the net worth would need to rise $40,000 a year. Increased revenues will be needed to compensate for inflation and to maintain and increase net worth.

Under action taken by the Executive Committee on March 8, 1974, the dues of regular members increased from $21 to $23 on January 1, 1975, with other dues and subscriptions rising in the same proportion. In view of the expected deficits for 1974 and 1975, the Executive Committee on December 27, 1974, voted to increase the dues further to $25 for regular members (and proportionate increases for subscriptions and other classes of membership) effective January 1, 1976. These actions were taken in accordance with a provision of the bylaws permitting the Executive Committee to raise dues in proportion to the rise in relevant wage and price indexes, provided that the increase not exceed 10 percent in any one year. They will not be sufficient to make the financial position of the Association sound unless the stock market makes an astonishing recovery. Under the circumstances, either further increases in dues, requiring a change in the bylaws approved by vote of the members, or significant cuts in expenditures to the detriment of the services rendered to the members will be necessary.

RENDIGS FELS, *Treasurer*

REPORT OF THE FINANCE COMMITTEE*

The accompanying inventory lists the securities held by the American Economic Association as of December 31, 1974, with costs and market values as of that date. The total market value of the securities portfolio at year end was $426,186. After making adjustments for cash additions and withdrawals, we estimate that the Association's investment portfolio (on a total return basis) declined in value by 28.5 percent during 1974. In comparison, although the widely quoted market indices were down 24 percent to 27 percent on the same basis last year, a more representative picture of the extent of the past year's market decline is provided by an unweighted average of all of the stocks on the New York Stock Exchange which was down 33.1 percent** in 1974. On a longer term basis, from the end of 1967 through December 31, 1974, which is a period encompassing contracting and expanding economic environments and up and down market cycles, the account's total return was +2.7 percent, while the Dow Jones Industrial Average, using comparable total return figures, was off 10.6 percent and the unweighted average of New York Stock Exchange stocks dropped 50.1 percent.**

It should be remembered that the $426,186 figure referred to above includes a Special Grant that was made by the Ford Foundation in January of 1969 and subsequently commingled with the Association's account. As of December 31, 1974, the Association's portion of the aggregate account was $298,143, or 70 percent, and the Special Grant represented the remaining $128,043, or 30 percent of the total.

During the past twelve months, several changes were made in the equity component of the account. This involved the sale of 500 Houston Lighting and Power, 1,500 Westinghouse Electric, 700 Newmont Mining, 300 Chase Manhattan Corporation, 1,000 Hawaii Bancorporation, 1,200 Kewanee Oil, and $15,000 par value Husky Oil convertible bonds; and new purchases of 600 Citicorp, and 1,200 Gulf Oil Corporation while additions in the following amounts were made to existing holdings: 100 Eastman Kodak, 200 First Bank Systems, 100 Merck, and 100 Xerox. The portfolio restructuring process carried out over the past two years has also had the effect of increasing the investment income of the portfolio by $5,335 or 43 percent.

Looking over the year just concluded, it

TABLE 1—INVENTORY SUMMARY AS OF DECEMBER 31, 1974

	Value	Percent	Estimated Income
Cash and Short-Term Securities	$ 35,012	8.2	$ 3,060
Medium-Term Securities	0	0.0	0
Long-Term Bonds and Preferred Stocks	0	0.0	0
Convertible Securities	0	0.0	0
Equity Securities	391,174	91.8	14,788
Total	426,186	100.0	17,848

---

* The Report of the Finance Committee is informational and is not an audited financial statement. Consequently, there may be some discrepancies between figures in the Report of the Finance Committee and the Auditors' Report which follows.

** An unweighted average of all the common stocks on the New York Stock Exchange adjusted to include the yield on the Standard & Poor's 500 Index.

TABLE 2—INVENTORY AND APPRAISAL AS OF DECEMBER 31, 1974

	Amount	Price	Value	Unit Cost	Total Cost	Est. Income
CASH AND SHORT-TERM SECURITIES (8.2 percent)						
Cash			($988)		($988)	
Secondary Reserves						
GMAC Nts (8.50 percent, 01/20/75)	36,000	$100	36,000	$100	36,000	$3,060
TOTAL CASH AND FIXED INCOME SECURITIES			35,012		35,012	3,060
EQUITY SECURITIES (91.8 percent)						
Utilities (3.8 percent)						
Central and Southwest	1,000	15	15,000	8	8,316*	1,120
Banks (15.5 percent)						
Citicorp	600	28	16,800	30	17,818*	480
First Bank System	700	33	23,100	20	13,831	1,015
Manufacturers Hanover Corp.	800	26	20,800	37	29,763*	1,376
			60,700		61,412	2,871
Other Financial (7.6 percent)						
Alexander and Alexander	900	26	23,400	19	16,725*	630
ERC Corp.	400	16	6,400	41	16,450	224
			29,800		33,175	854
Paper and Forest Products (2.0 percent)						
Hoerner Waldorf	1,000	8	8,000	12	12,107	720
Mining and Metals (4.0 percent)						
Utah International Inc.	400	39	15,600	25	10,088	340
Oil and Gas (15.9 percent)						
Atlantic Richfield	200	91	18,200	48	9,684	500
Continental Oil	500	45	22,500	33	16,448	900
Gulf Oil	1,200	18	21,600	17	19,981	2,040
			62,300		46,113	3,440
Drugs and Medical (5.1 percent)						
Merck	300	66	19,800	52	15,631*	420
Computers (8.1 percent)						
IBM	188	168	31,584	112	20,995*	1,128
Other Office Equipment (8.2 percent)						
Moore Corp.	300	41	12,300	37	10,950	360
Rank Organisation	2,200	2	4,400	10	22,513*	216
Xerox	300	51	15,300	62	18,608*	300
			32,000		52,071	876
Miscellaneous (29.8 percent)						
CBS	700	31	21,700	30	20,973*	1,022
Disney (Walt) Productions	416	21	8,736	3	1,079	50
Eastman Kodak	300	63	18,900	111	33,275*	618
Mapco	1,000	28	28,000	18	17,710	460
Minnesota Mining and Mfg.	300	46	13,800	81	24,448*	375
SR&F Cap Opportunities Fund	4,493	6	25,254	12	53,226*	494
			116,390		150,711	3,019
TOTAL EQUITY SECURITIES (100 percent)			391,174		410,619	14,788
TOTAL SECURITIES AND CASH			426,186		445,631	17,848

* More than one cost basis.

is important to note that there were an unusual number of developments, most of which were unprecedented and largely unpredictable. The most damaging of these events were the Arab oil embargo and the resulting high price of oil, the political turmoil culminating in the resignation of the Vice-President and President, the financial strains which appeared in the international monetary community, and the adverse weather conditions in the farm belt that led to disappointing crops and higher food prices. As a result of these difficulties, interest rates (which were also affected by the government's restrictive monetary policy) and the pace of inflation rose to unprecedented peacetime levels.

These developments exerted downward pressure on the equity markets causing investor confidence to fall to an extremely low level and stocks to sell at their lowest valuation in more than thirty years. Although 1975 will clearly be a recessionary year, the managers of the Fund anticipate that this continued slowing in economic activity will bring with it lower rates of inflation and a more moderate level of short-term interest rates and that an economic recovery will begin late in 1975. They further believe that, with security prices relative to underlying book value, earnings and cash flow extremely depressed, many of the uncertainties which exist today have been discounted and that any positive developments, such as those alluded to above in the areas of inflation and interest rates, could generate an upward revaluation of equities. However, the Fund managers do not expect investor confidence to be fully restored until meaningful progress is made with respect to the Middle East situation in general and a reduction in the price of oil in particular.

In view of this, we are persuaded that no change should be made at this time in the Committee's basic policy of maintaining a meaningful exposure to equities. However, recognizing that important questions—perhaps the most significant of which are the oil, political, and military aspects of the Middle East situation—are not yet fully resolved, we will continue to reevaluate the Association's investment position; and, should conditions warrant a change in the portfolio's composition, appropriate steps will be taken.

BERYL W. SPRINKEL, *Chairperson*

AUDITORS' REPORT

To the Executive Committee of
The American Economic Association:

We have examined the statements of assets and liabilities of THE AMERICAN ECONOMIC ASSOCIATION (a District of Columbia corporation, not for profit) as of December 31, 1974 and 1973, and the related statements of revenues and expenses, changes in general and restricted fund balances and changes in assets and liabilities for the years then ended. Our examination was made in accordance with generally accepted auditing standards, and accordingly included such tests of the accounting records and such other auditing procedures as we considered necessary in the circumstances.

In our opinion, the accompanying statements present fairly the assets and liabilities of The American Economic Association as of December 31, 1974 and 1973, and its revenues and expenses, changes in fund balances and the changes in its assets and liabilities for the years then ended, in conformity with generally accepted accounting principles consistently applied during the periods.

ARTHUR ANDERSEN & CO.

Memphis, Tennessee,
February 27, 1975.

THE AMERICAN ECONOMIC ASSOCIATION
STATEMENTS OF ASSETS AND LIABILITIES—DECEMBER 31, 1974 AND 1973

ASSETS	1974	1973	LIABILITIES AND FUND BALANCES	1974	1973
CASH, including $145,000 certificate of deposit in 1973	$ 37,471	$ 208,945	ACCOUNTS PAYABLE AND ACCRUED LIABILITIES	$252,710	$ 101,445
			DEFERRED INCOME (Note 1):		
INVESTMENTS, at market (cost of $848,753 and $655,688) (Notes 1 and 2)	829,174	776,023	Life membership dues	54,774	49,794
			Prepaid membership dues	215,298	205,209
			Prepaid subscriptions	148,961	133,802
RECEIVABLES, less allowance for doubtful accounts of $2,300 and $2,000	25,552	24,496		419,033	388,805
			ACCRUAL FOR DIRECTORY (Note 1)	—	80,000
			FUND BALANCES:		
			Restricted	307,552	313,502
PREPAID EXPENSES	1,448	2,043	Less—Unrecognized change in market value of investments (Notes 1 and 3)	(68,007)	(27,677)
				239,545	285,825
			General	148,381	226,610
			Less—Unrecognized change in market value of investments (Notes 1 and 3)	(157,931)	(64,272)
				(9,550)	162,338
OFFICE FURNITURE AND EQUIPMENT, at cost, less accumulated depreciation of $4,161 and $3,423	8,093	6,906	Total fund balances less unrecognized change in market value of investments	229,995	448,163
	$901,738	$1,018,413		$901,738	$1,018,413

The accompanying notes to financial statements are an integral part of these statements.

The American Economic Association
Statements of Revenues and Expenses
for the Years Ended December 31, 1974 and 1973

	1974	1973
Revenues from Dues and Activities:		
Membership dues	$350,064	$311,961
Subscriptions	219,299	194,526
Job Openings for Economists	4,487	—
Advertising	61,131	72,930
Sale of copies, republications and handbooks	21,727	17,765
Sale of mailing list	23,540	20,112
Annual meeting	17,427	7,491
Sundry	13,022	11,289
	710,697	636,074
Investment Gains (Losses) (Note 2)	(56,059)	1,135
Net revenues	654,638	637,209
Publication Expenses:		
American Economic Review	201,082	172,161
Journal of Economic Literature	249,359	221,450
Papers and Proceedings	51,625	45,042
Directory publication (Note 1)	68,780	40,000
Job Openings for Economists	16,472	—
	587,318	478,653
Operating and Administrative Expenses:		
General and administrative:		
Salaries	99,401	86,200
Rent	7,962	8,005
Other (Exhibit I)	60,450	49,321
Committee	13,220	16,550
Annual meeting	13,253	12,825
Interest and miscellaneous	1,827	1,039
	196,113	173,940
Total expenses	783,431	652,593
Revenues (Less) than Expenses	$(128,793)	$(15,384)

The accompanying notes to financial statements and Exhibit I are an integral part of these statements.

The American Economic Association
Statement of Changes in General Fund Balance
for the Years Ended December 31, 1974 and 1973

	Total	Operations	Market Value Adjustments
Balance at January 1, 1973	$203,344	$ 77,568	$125,776
Add—market value adjustments resulting from inflation (Note 1)	38,650	—	38,650
Deduct—expenses in excess of revenues	(15,384)	(15,384)	—
Balance at December 31, 1973	226,610	62,184	164,426
Add—market value adjustments resulting from inflation (Note 1)	50,564	—	50,564
Deduct—expenses in excess of revenues	(128,793)	(128,793)	—
Balance at December 31, 1974	$148,381	$(66,609)	$214,990

The accompanying notes to financial statements are an integral part of this statement.

The American Economic Association
Statement of Changes in Restricted Fund Balances
for the Year Ended December 31, 1973

	Balance at January 1	Receipts	Disbursements	Allocation of Investment Gains (Note 4)	Balance at December 31
The Ford Foundation grants for—					
Economics Institute's orientation program for foreign graduate students of economics	$257,342	$24,594	$(16,230)	$14,795	$280,501
Committee on the Status of Women in the Economics Profession	15,571	—	(14,551)	—	1,020
The Alfred P. Sloan Foundation and Chase Manhattan Bank grants for increase of educational opportunities for minority students in economics	25,000	—	(1,097)	—	23,903
Funds reserved by the Association for publication of revised editions of *Graduate Study in Economics*, a guide originally published with funds from a Ford Foundation grant	3,240	2,792	—	—	6,032
The Asia Foundation grant for Asian economists' membership dues to The American Economic Association and related travel expenses	10	4,149	(2,524)	—	1,635
Sundry	311	100	—	—	411
	$301,474	$31,635	$(34,402)	$14,795	$313,502

The accompanying notes to financial statements are an integral part of this statement.

The American Economic Association
Statement of Changes in Restricted Fund Balances
for the Year Ended December 31, 1974

	Balance at January 1	Receipts	Disbursements	Allocation of Investment (Losses) (Note 4)	Balance at December 31
The Ford Foundation grants for—					
Economics Institute's orientation program for foreign graduate students of economics....................	$280,501	$ 28,710	$ (42,086)	$(11,026)	$256,099
Committee on the Status of Women in the Economics Profession......	1,020	15,000	(16,020)	—	—
The Alfred P. Sloan Foundation, Chase Manhattan Bank and Ford Foundation grants for increase of educational opportunities for minority students in economics......................	23,903	57,507	(58,648)	—	22,762
Funds reserved by the Association for publication of revised editions of *Graduate Study in Economics*, a guide originally published with funds from a Ford Foundation grant...........	6,032	62	—	—	6,094
The Asia Foundation grant for Asian economists' membership dues to The American Economic Association and related travel expenses............	1,635	2,521	(3,741)	—	415
The Carnegie Foundation grant for the Committee on the Status of Women in the Economics Profession........	—	21,610	(689)	—	20,921
The Kazanjian Foundation grant for the Committee on Economic Education..	—	750	—	—	750
Sundry............................	411	100	—	—	511
	$313,502	$126,260	$(121,184)	$(11,026)	$307,552

The accompanying notes to financial statements are an integral part of this statement.

THE AMERICAN ECONOMIC ASSOCIATION
STATEMENTS OF CHANGES IN ASSETS AND LIABILITIES
FOR THE YEARS ENDED DECEMBER 31, 1974 AND 1973

	1974	1973
SOURCE (USE) OF FUNDS:		
Revenues (less) than expenses	$(128,793)	$(15,384)
Add—		
Realized gains (losses) from sales of investments	(50,572)	1,619
Noncash charges:		
Market value adjustments (Note 1)	90,055	18,821
Depreciation	737	800
Directory publication	—	40,000
Funds provided (used) by operations	(88,573)	45,856
(Increase) decrease in—		
Investments, at cost:		
Sales	346,186	394,135
Purchases	(539,251)	(483,885)
Receivables and prepaid expenses	(461)	137
Office furniture and equipment	(1,924)	(693)
Increase (decrease) in—		
Accounts payable and accrued liabilities	151,265	4,917
Deferred income	30,228	127,442
Accrual for directory	(80,000)	—
Restricted funds	11,056	3,489
Increase (decrease) in cash	$(171,474)	$ 91,398

The accompanying notes to financial statements are an integral part of these statements.

THE AMERICAN ECONOMIC ASSOCIATION
NOTES TO FINANCIAL STATEMENTS

(1) Significant Accounting Policies:

Investments—

The Association accounts for its investments on a market value basis. Under the method used by the Association to value investments, the change in market value of corporate stocks during the year, after adjusting for an inflation factor (11.8 percent in 1974 and 7.0 percent in 1973), is recognized in income over a three-year period. The change in market value of corporate bonds and obligations, after adjusting for the inflation factor, is reflected in income currently. The changes in market value of investments are allocated to the general and restricted fund balances as appropriate.

Accrual for Directory—

Approximately every three to five years, the Association publishes a Directory which lists, among other things, the names and addresses of its membership. This Directory was published in 1974 and distributed at no cost to the membership. In order to more properly match the publishing cost of this Directory with revenue from membership dues, the Association has provided annually a reserve for estimated publishing costs. Actual costs of $68,780 in excess of accrued amounts at the beginning of the year were expensed currently.

Deferred Income—

Revenue from membership dues and subscriptions to the various periodicals of the Association are deferred when received, with these amounts then recognized as income quarterly as publications are mailed to the members and subscribers.

Life membership dues are also deferred when received, with subsequent recognition as income over the estimated average life of these members.

(2) Investments and Investment Income:

The following is a summary of investments held by the Association at December 31:

	1974		1973	
	Cost	Market	Cost	Market
Corporate bonds and obligations	$438,000	$438,000	$209,750	$210,500
Corporate stocks	410,753	391,174	445,938	565,523
	$848,753	$829,174	$655,688	$776,023

Investment gains (losses) recognized in income for the years ended December 31 were as follows:

	1974	1973
CORPORATE BONDS AND OBLIGATIONS—		
Interest	$ 23,720	$13,270
Decline in market value	(11,680)	(6,649)
	12,040	6,621
CORPORATE STOCKS—		
Cash dividends	16,256	12,943
Decline in market value recognized (Note 3)	(117,154)	(20,277)
	(100,898)	(7,334)
OTHER—		
Investment losses allocated to a restricted fund (Note 4)	32,799	1,848
INVESTMENT GAINS (LOSSES) INCLUDED IN INCOME	$ (56,059)	$ 1,135

(3) Unrecognized Change in Market Value of Investments:

As described more fully in Note (1), the Association recognizes in income over a three-year period changes in the market value of its corporate stocks. The following summarizes the years in which market value changes in stocks occurred that affect 1973 and 1974 revenues, and the amount of these market value increases (declines) that will be recognized in income in future periods.

Year of Market Value Change	Recognized In Income In		To Be Recognized In		Unrecognized Change as of December 31	
	1973	1974	1975	1976	1974	1973
1971	$ 13,163	$ —	$ —	$ —	$ —	$ —
1972	25,069	25,069	—	—	—	25,069
1973	(58,509)	(58,509)	(58,509)	—	(58,509)	(117,018)
1974	—	(83,714)	(83,714)	(83,715)	(167,429)	—
	$(20,277)	$(117,154)	$(142,223)	$(83,715)	$(225,938)	$ (91,949)

Included in the above unrecognized changes as of December 31 are declines of $68,007 and $27,677 in 1974 and 1973, respectively, which have been allocated to a restricted fund. The amounts allocated are based on the percentage of the Association's total stock portfolio owned by this restricted fund.

(4) Restricted Fund:

In 1968, the Association entered into an agreement with the University of Colorado relating to the Ford Foundation grant for the Economics Institute which provides, among other things, that the Association invest a portion of the funds received and allocate any income and market value adjustments therefrom to the restricted fund. In accordance with this agreement, the following adjustments were allocated to the restricted fund:

	1974	1973
Net investment (losses) (Note 2)	$(32,799)	$(1,848)
Market value adjustments arising from inflation	21,773	16,643
	$(11,026)	$14,795

(5) Retirement Annuity Plan:

Employees of the Association are eligible for participation in a contributory retirement annuity plan. Payments by the Association and participating employees are based on the employee's compensation. Benefit payments are based on the amounts accumulated from such contributions. The total pension expense was $6,798 and $5,355 for 1974 and 1973, respectively.

(6) The Association:

The American Economic Association is an educational organization, substantially exempt from income tax under section 501(c)(3) of the U.S. Internal Revenue Code. As required by Section 511(a) of this Code, the Association provides for Federal income taxes on certain revenues which are not substantially related to its tax exempt purpose. This "unrelated business income" includes income from advertising and the sale of mailing lists.

The Association has been determined to be an organization which is not a private foundation.

EXHIBIT I—THE AMERICAN ECONOMIC ASSOCIATION
STATEMENTS OF OTHER GENERAL AND ADMINISTRATIVE EXPENSES
FOR THE YEARS ENDED DECEMBER 31, 1974 AND 1973

	1974	1973
Mailing list file maintenance and periodic mailing expenses	$13,598	$14,205
Accounting and legal	7,500	7,300
Office supplies	8,998	5,507
Postage	11,475	5,383
Dues and subscriptions	2,800	3,325
Telephone	3,130	3,063
Provision for Federal income taxes (Note 6)	3,371	3,000
Investment counsel and custodian fees	2,783	2,865
President and president-elect expenses	4,200	2,334
Travel and entertainment	1,327	1,084
Depreciation (straight-line method)	737	800
Uncollectible receivables	586	262
Currency exchange charges (credits)	(55)	193
	$60,450	$49,321

REPORT OF THE MANAGING EDITOR AMERICAN ECONOMIC REVIEW

The number of manuscripts submitted in 1974 was 723, slightly less than last year, and approximately the same as in 1972. We printed 125 papers in all, 14 more than last year. The comparative statistics for the last twenty years are shown in Table 1.

Manuscripts Submitted, Accepted, and Published

The flow of manuscript submissions appears to be at a plateau of 700 to 800 a year, a level reached in 1969. This is double the number submitted ten years ago, and roughly three times the level of the mid-1950's. The present format of the *Review* easily permits publication of 110 to 140 papers a year, which also doubles the publication rate reached ten years ago. Nevertheless, we continue to publish only 16 to 20 percent of the papers submitted. The high rejection rate is not due to limitations imposed by the format. It springs instead from the combined result of refereeing and editorial decisions. I suspect that many rejected papers are reworked and ultimately published elsewhere. From conversations with other journal editors and from my reading of other journals, I do know that we publish each other's rejected papers. A potential author benefits from the diversity of tastes and perceptions exercised by the different journals.

TABLE 1—MANUSCRIPTS SUBMITTED AND PUBLISHED, 1954–74

Year	Submitted	Published	Ratio of Published to Submitted
1954	231	47	.20
1955	245	41	.17
1956	242	48	.20
1957	215	40	.19
1958	242	46	.19
1959	279	48	.17
1960	276	46	.17
1961	305	47	.15
1962	273	46	.17
1963	329	46	.14
1964	431	67	.16
1965	420	59	.14
1966	451	62	.14
1967	534	94	.18
1968	637	93	.15
1969	758	121	.16
1970	879	120	.14
1971	813	115	.14
1972	714	143	.20
1973	758	111	.15
1974	723	125	.17

The backlog of accepted papers is the same as last year. As of November 30, there are 81 accepted papers which have not yet been published. Twenty-nine will appear in the March 1975 issue, the remaining 52 in June or September 1975.

We printed 125 papers this year, 49 main articles and 76 shorter papers. The latter include short articles, notes, comments, replies, and special obituaries. The special obituaries are written at my request to take note of the role which a number of individuals have played in the history of economic analysis.

As shown in Table 2, the size of the *Review* has been kept to 1,146 pages, the same as last year. The size of the December issue was kept under control by a reduction in the number of pages devoted to the Ph.D. dissertations, from 67 last year to 25 in 1974. The reduction was made possible by the elimination of the 75-word abstract. We now print only the author's name, institution, year, and title, classified by field. The abstracts are available in a publication of University Microfilms, which most university reference libraries can obtain. It seemed to be a waste of

TABLE 2—SUMMARY OF CONTENTS, 1973 AND 1974

	1973		1974	
	Number	Pages	Number	Pages
Articles	54	747	49	684
Shorter Papers, including Notes, Comments, and Replies	57	273	73	386
Special Articles	2	15	3	4
Dissertations		67		25
Announcements and Notes		44		47
Total	113	1,146	125	1,146

space to publish them in the *Review* as well. The *Review* will continue to publish the annual dissertation list.

Subject Matter of Submitted and Printed Manuscripts

Table 3 shows the subject matter distribution of submitted and printed manuscripts. The most popular fields are microeconomics, macro and monetary theory, international economics, labor economics, and welfare economics. These six fields comprise 60 percent of the papers submitted and papers published.

I have continued the practice of inviting papers on subject matter of special interest. In 1973 and 1974, the *Review* published the Nobel Prize addresses by Simon Kuznets, Kenneth Arrow, and Wassily Leontief. In 1973, we published an invited paper by Don Patinkin on the work and teaching of Frank Knight. In 1974, we published four invited reviews of the annual report of the Council of Economic Advisors (*CEA*) and an invited review of the 60th Annual Report of the Board of Governors of the Federal Reserve System. In 1975, we will continue to publish re-

TABLE 3—SUBJECT MATTER DISTRIBUTION OF SUBMITTED AND PUBLISHED MANUSCRIPTS IN 1974

	Submitted	Published
General Economics and General Equilibrium Theory	15	4
Microeconomic Theory	115	22
Macroeconomic Theory	41	6
Welfare Theory	65	18
Economic History, History of Thought, Methodology	13	2
Economic Systems	10	1
Economic Growth, Development, Planning, Fluctuations	37	11
Economic Statistics	17	5
Monetary and Financial Theory and Institutions	85	7
Fiscal Policy and Public Finance	27	—
International Economics	61	12
Administration, Business Finance	34	6
Industrial Organization	41	6
Agriculture, Natural Resources	33	8
Manpower, Labor, Population	94	10
Welfare Programs, Consumer Economics, Urban and Regional Economics	35	7
Total	723	125

TABLE 4—COPIES PRINTED, SIZE, AND COST OF PRINTING AND MAILING IN 1974

	Number of Copies Printed	Pages		Cost		
		Net	Gross	Issue[a]	Reprints[b]	Total
March	27,500	252	312	$ 28,755.70	$ 496.88	$ 29,252.58
June	28,500	270	304	30,604.92	481.35	31,086.27
September	27,500	300	332	36,389.44	467.15	36,856.59
December	27,000	324	384	41,856.08[c]	450.00[c]	42,306.08[c]
Total	110,500	1,146	1,332	$137,606.14[c]	$1,895.38[c]	$139,501.52[c]

[a] Includes allocated cost of preparing mailing list.

[b] Costs are reduced by a credit resulting from charges to authors for additional reprints.

[c] Estimate.

views of the *CEA* report, but will invite a smaller number.

Anonymous Refereeing

Beginning January 1974, all papers have been refereed anonymously. Authors are requested to omit their name and institutional affiliation from the title page and title footnote. The practice was discussed in my report last year. It was adopted to make certain that knowledge of the author's identity or employer's identity would play no role in evaluation of the paper.

Only one referee out of some 400 has objected to the practice and refused to read a paper because the author was not identified. No other objections have occurred. If referees comment at all on the practice, it is favorable. Occasionally a referee requests the name of the author after the evaluation is carried out. We will provide that information. A referee may know the author's identity, either through prior reading of the paper, or through repeated self-referencing in the text. So far I don't have the impression that self-referencing is carried out for the purpose of piercing the veil of anonymity.

Expenses—Printing and Mailing

The 1974 costs of printing and mailing the *Review* are shown in Table 4 and for other years in Table 5. The net number of pages includes only those pages listed in our table of contents plus the index. Advertising pages then make up the difference between gross and net. The 1974

TABLE 5—ACTUAL AND BUDGETED EXPENDITURES, 1967–1975

	Printing and Mailing	Payments to Contributors	Office Expenses	Total
1967	$ 82,840	$2,255	$27,114	$112,209
1968	92,948	2,140	29,155	124,243
1969	97,183	600	26,765	124,548
1970	111,227		36,336	147,564
1971	120,120		43,524	163,644
1972	107,196		44,473	151,669
1973	117,873		49,121	166,994
1974 (budget)	121,518		59,749	181,267
1974 (actual)[a]	139,502		58,396	197,898
1975 (budget)	165,742		65,911	231,653

[a] Estimate.

printing expenses exceed the estimate I made last December for a number of reasons: in 1974, 4,500 more copies were printed than in 1973; mailing costs rose; a paper surcharge has been added to the regular printing contract to reflect the increased price of this material; and an across-the-board increase in other printing charges took effect with the September issue. In all, our 1974 printing and mailing costs will come to $139,501, an excess of $18,000 over the budgeted figure for this year. I expect printing costs to rise again in 1975. There are two reasons: the increases in the printing contract and in mailing charges have not been in effect for one full year. Moreover, another increase in printing costs may occur and it would be desirable to appropriate funds for this contingency. Of course, the current business recession may permit us (in the short run) to withstand a request for an increase by taking advantage of the greater degree of aggressive rivalry among printing companies.

Charges to Authors

Continued increases in costs of printing are a matter of concern because they are eventually reflected in higher membership dues. Other ways to defray printing costs require a direct levy on the author or prospective author. Some journals have used a page charge to reduce the share of the cost falling on the readers. This practice would appear most practicable in those fields where research receives outside financial support and the page charge ultimately comes out of a research contract or grant. Research support in economics may not be sufficiently widespread to allow page charges. But it should not be ruled out for future consideration. Another type of levy on the author is the submission fee. At the moment, the *Review* charges a ten dollar fee. The proceeds are used to pay for manuscript screening. This was described in my annual report for the year 1971. Other journals in economics also use submission fees, and some may devote the proceeds to production costs. This practice could become more widespread in the future.

Office Expense

Office expenses rose in 1974 because of higher salaries, increased supply costs, higher fringe benefits, and additional editorial assistance. The 1974 figure shown in Table 5 is still an estimate as all charges are not yet in. Further increases in office expenses will occur in 1975 as a consequence of increased personnel and material costs.

Board of Editors

Five members of the Board of Editors complete their terms at this time: Robert Baldwin, Jack Hirschleifer, Sherwin Rosen, Thomas Saving, and Anna Schwartz. I wish to thank them for their hard work, cooperation, and critical standards.

I am submitting to the Executive Committee the names of four nominees to serve three-year terms until December 31, 1977. They are: Eugene Fama, Robert J. Gordon, James Melvin, and William Nordhaus. Anna Schwartz has graciously accepted a second term.

I should like to express my thanks to the continuing members of the Board of Editors: Bela Balassa, Martin Feldstein, Ann Friedlaender, Stephen Goldfeld, Robert Hall, Bent Hansen, Leonard Rapping, Stephen Resnick, Jerome Stein, Joseph Stiglitz, S. C. Tsiang, Finis Welch, and Marina Whitman. I appreciate their work and assistance.

The following assisted me this year as editorial consultants: Raveendra Batra, Ernst Berndt, Roger Betancourt, Alan Blinder, Byron Brown, Allan Feldman, Donald Frey, Paul Ginsburg, Gerald Goldstein, Michael Grossman, William Haley, James A. Hanson, Jerry Hausman, John Kennan, Rachel McCulloch, James Ohls, Mark Pauly, J. David Richardson, John Roberts, Robert Rohr, José Alexandre Scheinkman, John Taylor, and Warren Weber.

I have received valuable proofreading and mathematical assistance from: Clemens F. J. Boonekamp, Kenneth Chang, Robert Jones, Chon Pyo Lee, Hiroshi Ono, and Orlando Roncesvalles. Finally, I am grateful to Wilma St. John for her fine work as assistant editor, and to Carol Chapin, editorial assistant.

In addition to the members of the Board and the editorial consultants, I have sought and received the assistance of a large number of economists during the course of the year. I wish to thank them for their cooperation and application of high standards in reading and evaluating manuscripts. They have eased the work load that would have otherwise fallen on the Board of Editors. The following have assisted in this way:

H. Aaron
I. Adelman
D. K. Adie
B. Aghevli
N. Aitken
A. Alchian
P. Allen
R. C. Amacher
J. Anderson
F. D. Arditti
S. W. Arndt
O. Ashenfelter
E. Bailey
M. Bailey
M. Balch
E. Baltensperger
V. N. Bandera
R. Barlow
D. P. Baron
N. S. Barrett
Y. Barzel
R. N. Batra
W. J. Baumol
M. Beckmann
F. Bell
G. Benston
U. Ben-Zion
A. Bergson
T. Bergstrom
T. J. Bertrand
S. Besen
K. Bhatia
G. O. Bierwag
F. Black
A. Blinder
M. Blume
D. R. Bohi
L. A. Boland
T. E. Borcherding
M. Boskin
R. Bower
D. Bradford
W. Brainard
W. H. Branson
F. P. R. Brechling
G. Brown, Jr.
J. Brown
E. K. Browning
H. J. Bruton
E. Burmeister
P. Cagan
J. A. Carlson
S. L. Carroll
F. Casas
M. Chacholiades
W. Chang
P. L. Cheng
S. Cheung
B. R. Chiswick
G. C. Chow
C. Christ
C. Cicchetti
C. S. Clark
S. Clemhout
J. Conlisk
M. Connolly
J. C. Cragg
J. Crotty
R. G. Cummings
M. R. Darby
R. H. Day
R. Deacon
L. De Alessi
M. De Groot
D. De Tray
W. E. Diewert
E. Dolan
F. T. Dolbear, Jr.
R. Dorfman
R. Dornbusch
G. Douglas
D. J. Dugan
W. C. Dunkelberg
G. Eads
R. A. Easterlin
N. Edelson
R. Ehrenberg
R. Eisner
B. Ellickson
J. W. Elliott
S. Engerman
E. Fama
E. L. Feige
A. Feldman
S. Fischer
A. Fisher
R. Flanagan
J. Flanders
F. Flatters
B. Fleisher
J. E. Floyd
M. Freeman
R. Freeman
J. Fried
B. Friedman
I. Friend
G. Fromm
M. Galatin
L. E. Gallaway
D. Gaskins, Jr.
S. K. Ghosh
W. E. Gibson
A. Gifford
L. Girton
C. Goetz
R. Goldfarb
M. J. Gordon
R. J. Gordon
M. Gort
E. Gramlich
W. L. Gramm
W. P. Gramm
J. Griffin
H. Grossman
M. Grossman
J. G. Gurley
J. Gwartney
J. Hadar
R. Hamada
M. Hamburger
D. Hamermesh
O. Hart
J. C. Hause
R. Haveman
G. A. Hay
J. Heckman
P. Heller
P. Helmberger
J. V. Henderson
W. Hettich
J. Hixson
H. Hochman
D. Hodgman
C. C. Holt

H. Hori
D. Jaffee
E. James
F. Jen
M. B. Johnson
G. Johnson
T. Johnson
E. B. Jones
D. W. Jorgenson
D. A. Katz
G. G. Kaufman
D. Keesing
M. G. Kelly
J. Kennan
R. A. Kessel
C. Khang
J. Kindahl
K. Klappholz
B. Klein
A. K. Klevorick
J. Kmenta
T. Koizumi
R. Korkie
M. B. Krauss
M. E. Kreinin
R. Kuller
P. Kumar
M. Kurz
K. J. Lancaster
W. Landes
L. J. Lau
L. B. Lave
S. Lebergott
T. H. Lee
A. Leibowitz
H. Levin
H. Levy
C. M. Lindsay
R. Lucas
D. G. Luckett
J. J. McCall
B. T. McCallum
T. McCoy
R. McCulloch
D. McFadden
D. L. McNicol
S. P. Magee
B. Malkiel
A. Manne
E. Mansfield
R. T. Masson
S. Masters
J. P. Mattila
C. Maurice
A. H. Meltzer
J. R. Melvin
C. Metcalf
P. Meyer
R. T. Michael
P. Mieszkowski
L. S. Miller
E. S. Mills
H. Minsky
H. Mohring
S. Morley
D. Mueller
M. I. Nadiri
K. Nagatani
P. Neher
C. Nelson
R. Nelson
M. Nerlove
J. Newhouse
P. K. Newman
D. Nichols
J. Niehans
R. Noll
W. Nordhaus
W. Oakland
W. Oi
A. Okun
E. O. Olsen
M. Olson
J. M. Ostroy
R. Parks
D. Parsons
M. Parvin
M. V. Pauly
S. Peltzman
J. Pencavel
R. G. Penner
G. Perry
W. Peterson
E. S. Phelps
R. S. Pindyck
T. Pinfold
M. Piore
C. G. Plourde
R. Pollak
W. Poole
R. B. Porter
R. D. Portes
D. Purvis
R. E. Quandt
W. Ramm
A. Razin
U. E. Reinhardt
T. Ribich
G. Rich
M. J. Ringo
L. S. Ritter
R. J. Rohr
R. Roll
D. Roper
S. Ross
J. Rosse
S. Rottenberg
J. C. R. Rowley
H. Ryder, Jr.
D. Salkever
A. Sandmo
A. Santomero
T. Sargent
R. Sato
F. M. Scherer
R. Schmalensee
A. Schmitz
R. Schuler
M. Schupack
N. L. Schwartz
G. W. Scully
J. B. Sheahan
W. G. Shepherd
C. D. Siebert
W. Silber
E. Silberberg
N. J. Simler
C. A. Sims
D. Sjoquist
V. Smith
L. Solmon
R. M. Solow
H. F. Sonnenschein
C. Southey
F. Stafford
R. M. Starr
P. Steiner
N. Stern
G. Stevens
W. J. Stull
M. Taussig
E. A. Thompson
J. Thornton
A. W. Throop
T. N. Tideman
G. Tolley
R. D. Tollison
W. Travis
S. Turnovsky
D. Vandermeulen
E. C. H. Veendorp
W. S. Vickrey
M. Visscher
R. Vogel
G. Walton
H. Wan
H. W. Watts
R. Waud
R. S. Weckstein
R. L. Weil
L. W. Weiss
S. H. Wellisz
F. W. Westfield
G. A. Whitmore
R. Willig
R. J. Willis
R. Wilson
D. Wisecarver
D. A. Worcester, Jr.
S. Y. Wu
L. B. Yeager
Y-H. Yeh
E. Zabel
R. Zecher
R. Zeckhauser

GEORGE H. BORTS, *Managing Editor*

REPORT OF THE MANAGING EDITOR JOURNAL OF ECONOMIC LITERATURE

This year represents the end of my second term as managing editor. Accordingly, it seems appropriate to sketch what my staff and I have undertaken during these past six years, to discuss what has happened in 1974, and to suggest some of the things with which we are concerned for 1975.

Six-year Review

When the *Journal of Economic Literature* was established, the objective was to have a journal primarily concerned with the literature in economics. The plan was to divide the effort basically in two parts:

I. There was to be a limited number of articles on emergent trends in the literature. These included:
   (1) Comprehensive surveys of literature developed within the prior decade or so in subfields.
   (2) Interpretive essays on subfields.
   (3) Miscellaneous elements such as articles on trends in literature, etc.

II. There was to be a major bibliographic effort, including:
   (1) The absorption of the new book listing and reviewing function of the *American Economic Review*. This absorption involved a changing from listing books, as in the *AER*, to providing brief annotations and classification of them in accordance with the American Economic Association 2-digit classification system.
   (2) The absorption of the *Journal of Economic Abstracts* and of the *Index of Economic Articles*. In this connection we were directed to:
      (*a*) Use in the *Journal of Economic Literature* the 2- and 3-digit classification system developed by the Classification Committee of the American Economic Association in 1967.
      (b) Publish abstracts.
      (c) Develop a compatible fourth digit to be used for classification for the annual *Index of Economic Articles*.
      (d) Prepare the annual *Index of Economic Articles*.

Articles

Our plan has been to publish four survey articles each year. In some instances we have published only three or even two, either because articles which had been assigned did not appear in publishable form or because, as in 1974, there was an opportunity to publish a commemorative article. That 1974 opportunity seemed to serve a double purpose; not only did it result in what is clearly a significant contribution itself, but it also drew attention to an earlier article of particular historical importance. We have also published a variety of articles on the literature, some review articles on particularly noteworthy books, and several articles about the development of economic literature, as well as some communications which are largely an outgrowth of reactions to our survey articles.

Books

We have annotated 1,200–1,300 books per year, classified in thirty-six categories, as well as all new journals (or older ones not previously covered). Reviews of about 170–185 books per year are published.

Journals

Journal Selection

The directive to absorb the *Journal of Economic Abstracts* and continue *Index of Economic Abstracts* meant that there was no option in the initial selection of journals. The records sent us from John Gurley's *American Economic Review* office, which maintained the list, showed that it received 254 journals for listing purposes in the *Journal of Economic Abstracts*. In addition, there were another thirty-five journals included in the *Index of Economic Abstracts* but not listed in the *Journal of Economic Abstracts*. The annual *Index* was to be compiled from our quarterly data and from volumes of collected essays and Congressional documents.

As time has passed, we have reduced the number of these old, inherited agreements in order to make space for new or other journals, as well as eliminated some which seemed to be inappropriate. In our six years we have tried to maintain a reasonable relationship between dropping earlier listings and abstracting agreements and adding new ones. You will note, later on when we come to Table 1, that we have listed the contents (in whole or in part) of 986 journal issues (drawn from 219 journals) in the four issues published during 1974. This shift represents a reduction from the 1,011 journals handled in 1973, but it is also an increase of some significance over the number of journals we handled in the first three years of our existence.[1] The plain fact that journals are proliferating in number poses a burden on our earlier decision to list and abstract where we thought appropriate.

Those journals which are listed but are only in part economics in content are selectively listed. We have also ceased to classify articles in foreign languages in the absence of an English summary or abstract because of the difficulties of finding personnel.

Abstracts

In the interest of cost reduction we have instituted two changes in the abstracting arrangements. We have reduced the maximum length of abstracts from 300 to 100 words; an interesting byproduct has been what seems to us an apparent improvement in quality and accuracy. We have also made most abstracting agreements selective, generally restricted to about one-half the articles in any annual output of a journal.

Article Classification

The application of the *AEA* classification system involved formulation of key and marginal concepts in each subject class. It also required the training and supervision of the requisite part-time personnel, generally graduate students. The decision to use part-time personnel was motivated by an attempt to hold costs down and to afford training to graduate students by intensive exposure to current literature.

Fourth Digit

We have also completed, insofar as one ever completes this sort of thing, the mandate to develop a fourth digit for the new *AEA-JEL* classification scheme and thereby increased the number of "pigeon-holes" from slightly under a hundred to somewhat more than three hundred. The fourth-digit material will be available as we publish the annual *Indexes*. Mrs. Truus Koopmans (Yale) and A. Maeshiro and J. Wells (both of the University of Pittsburgh) as well as the 1970 cadre of classifiers (Leigh Boske,

---

[1] To some extent the increase during the early years represents rationalization of procedures—making sure there were complete volumes for the journals covered; maintaining sequence order in journals; printing Tables of Contents and abstracts at the same time. The establishment of systematic listing thus delayed the inclusion of certain journals and probably led to abnormally low numbers processed in the first two years.

Thomas Mason, Manharlal Vyas, Andre Lumbroso, and Donald DePamphilis) were helpful in this effort.

Annual Indexes

As I shall note below, it has taken us somewhat longer to work out the publication of the annual *Indexes* than we had anticipated; but the 1970 volume is presently in the hands of the printer. It will be followed by the 1971, the 1972, and the 1969 volumes. Preparation of volumes has turned out to be far from the marginal job that I had once anticipated. Our task was probably complicated by a decision to introduce computerization of articles listings for immediate use in *JEL* and subsequent use in the *Indexes*. In any event, the additional work included the selection and classification of articles in volumes of collected essays and certain United States government documents. For 1970, the articles included from books represented an addition of about 40 percent of articles already classified for journals.

As I mentioned, we have indexed over 200 journals for each of the six years we have published the *Journal*. We have also indexed material in 200 books of collected essays, plus selected articles in several hundred documents for 1969, 1970, 1971, and 1972. Our proposal is to continue the indexing of books of collected essays for 1973 and 1974, but not to include the selected government documents. It turns out that the cost of each government document entry is very high (a multiple of ten to twenty times the cost of the other essays), and budget constraint identifies that as one of the leading candidates for cost cutting. We have prepared a corrected tape of journal articles, collected essays, and documents for 1970 and proofread that print-out; it is now at the printer. We have proofread the computer print-outs of the journal data for 1971 and 1972 and prepared inputting sheets of the books and documents for the Kingsport Press computer. We are on the threshold of preparing inputting sheets of the complete 1969 file; it will then be ready to send to Kingsport for inputting on the computer. The 1973 print-out of journal material is awaiting proofreading. Our hope is to take advantage of this momentum and bring the annual *Indexes* up to publication date about eighteen months after the end of the title year; for example, the 1974 volume should be ready by July of 1976. This apparently long delay is not completely a product of our making. A great many journals dated 1974 will not be published and in our hands until the autumn of 1975 (such are the delays associated with modern printing, to say nothing of modern postal service). If they are in our hands by September of 1975, we can assume that they will be classified, checked, and proofread by the end of 1975. This kind of timing means that the whole manuscript will be ready to go to the printer in the early months of 1976. I believe it is not unlikely that the printer should be allowed about six months to produce the finished product. That is how July of 1976 seems to be a reasonable target date.

What has taken so long in the preparation of these volumes has been computerizing of the data base. This data base, in the form originally suggested to us by Richard Ruggles, has been on various computers since the summer of 1971. Originally we used the University of Pittsburgh computer, but our base had to be transferred to Yale University when the University of Pittsburgh shifted from an *IBM* system to another computer system with a different computer language. We continued to use the Yale University computer, through the kindly intervention of and generous subvention by the National Bureau of Economic Research (this is a splendid opportunity to thank John R. Meyer and Orin Hansen, as well as to take notice of the aid we were given from a grant to the National Bureau of Economic Research by the International Business Machine Corporation). About a year ago we began the transfer of our data processing from the National Bureau of Economic Research facilities on the Yale University computer to Kingsport Press in Kingsport, Tennessee. This final transfer, which we assumed would not be too difficult, has proceeded with dispatch. Of course, some adjustments have had to be made for differences in software as well as compatibility of programs, but, all in all, we have made the transition with reasonably little pain.

Computerizing the data base, which in effect makes it possible to access data entries by journal (issue), by author, by any one of several subject matter codes (including geographical descriptors), or by year of publication, requires far more proofreading than we had even anticipated. My staff often speculates, when we see obvious errors because of letter substitution or items left out, whether "numbers crunchers" have the same troubles that people who print upper and lower case words have. In any event, computerizing of the data base was a "noble experiment"; unlike previous noble experiments, we have not turned our back on the experience, but we are somewhat the wiser about the costs involved.

Report on 1974

Table 1—Quantitative Analysis of Contents (1969–74) again illustrates our product. Members will note that we seem to have published a slightly smaller journal than we published in 1973. In reality, it actually contains more words because we shifted for 1974 to a more compact typeface, Times Roman. Several people have noted a difficulty in reading this compact typeface. We are planning to revert to a typeface comparable to the one we used prior to 1974, but this shift, it was decided,

TABLE 1—QUANTITATIVE ANALYSIS OF CONTENTS
Journal of Economic Literature 1969–74

	1969		*1970*		*1971*		*1972*		*1973*		*1974*	
	Number	Pages	Number	Pages	Number	Pages	Number	Pages	Number	Pages	Number	Pages
Survey articles	4	111	4	123	1	37	3	78	4	144	3	102
Essays on subfields	3	55	4	82	5	105	5	129	2	28	5	99
Review articles	4	47	6	39	5	40	5	39	—	—	1	5
Articles about economic literature	1	12	—	—	2	32	1	19	2	38	—	—
Communications	12	4	7	18	7	24	4	13	10	26	13	71
Books annotated	1,203	182	1,145	171	1,279	206	1,209	257	1,214	239	1,211	229
Books reviewed	188	243	177	259	172	249	160	241	175	259	168	239
Journal issues listed and indexed	668	101	717	125	835	130	849	140	1,011	185	986	180
Number of individual articles							5,387		7,218		7,360	
Subject index of journal articles	—	195	—	296	—	251	—	263	—	357	—	338
Abstracts of articles	1,139	444	1,181	356	1,426	337	1,415	313	1,906	407	1,645	312
Total pages[a]		1,432		1,539		1,479		1,572		1,748		1,718[b]

[a] Includes, in addition to listed pages, classification systems, table of contents, indices, journal subscription information, etc.

[b] Estimate.

would take place only with the next issue (March 1975).

As I noted above, we printed only three survey articles this year. They covered, however, 102 pages. We printed five essays in subfields which covered ninety-nine pages, but one of these essays on a subfield was a substitution for a survey article. We printed only one review article this year. There were no articles about economic literature, however. We had a record number of communications, largely mop-up operations and comments about previous articles. We annotated about the same number of books which we have done for the past few years. We reviewed 168 books; the reviews were longer and, we hope, more analytical. As I noted above, we indexed fewer journals' issues than in the previous year, but still a high number. We actually indexed a larger number of articles (7,360) than we had done in previous years. One decision which I took early in the year was to "catch up" and thus to increase the number of article listings in the early months of the 1974 printing contract because costs of printing this material was likely to go up as the year progressed. You will note that we have fewer pages devoted to a subject index of journal articles, but that again is a phenomenon associated with a more compact typeface. We had fewer abstracts of articles, reflecting a decision made by the Board of Editors two years ago to request editors of journals with which we had abstracting agreements to submit abstracts for only a portion of the articles they had printed.

My annual report in December 1973 included an explanation of the processes we use to train annotators and article classifiers. The Associate Editor supervises these processes as well as reviewing the articles, book reviews, and annotations with me for subject matter consistency. The Assistant Editor, Drucilla Ekwurzel, is an economist and, besides preparing the manuscript for the printer and seeing the proofs through the various stages, is able to help us with the subject matter monitoring.

The Chancellor and the Dean of the Faculty of Arts and Sciences of the University of Pittsburgh have helped the American Economic Association by the assignment of two to three teaching assistants to the *Journal* plus the allocation of space and equipment.

I have included in Table 2 brief reference to the costs of putting out the *Journal of Economic Literature*. It is appropriate that the cost problem be basically discussed most thoroughly with the Executive Committee, since it is for that reason, amongst others, that the responsibility for the administration of the Association is vested in

TABLE 2—EXPENDITURES 1969 THROUGH 1974[a]

	JEL Printing[b]	*JEL/Index* Editorial, Bibliographical, Payments to Contributors, etc.	Total
1969	$104,721	$47,087	$151,808
1970	116,845	63,544	181,090
1971	115,127	74,284	189,410
1972	118,442	85,083	203,525
1973	134,418	87,032	221,450
1974 Estimated Expenditures	139,400	93,155	232,555

[a] Figures for 1969–73 are coordinated with the auditors' figures.

[b] Includes mailing, addressing, and gross cost of reprints.

the Executive Committee. We have pursued what we believe are cost-effective policies. We anticipate an increase in the printing cost of putting out the *Journal*, particularly because of the rise of paper cost. However, the fact is that our costs have not risen in proportion to the rise of retail prices. Part of the reason for that moderation is a reluctance to raise the basic cost of hourly help. I have asked the Executive Committee to permit me to raise this price because the *Journal* is no longer competitive with other projects around the University of Pittsburgh. Thus, I am asking for an increase in our budget for 1975; however, it, too, is not commensurate with the rise in the cost of living. Part of this latter example of restraint can be explained by increased productivity of our staff (the turnover has been very light), but part of it also is explained by a posture of studied administrative cynicism.

Table 3 refers to the technical difficulty of material in the *Journal of Economic Literature*. Again, we have tried to maintain a viable balance between material that will interest subject matter specialists and material that will be of interest to all the members of the Association. The problem is not only one of technical jargon; it is also a problem of subject matter and analytical method. I am grateful to the number of members who have written in with suggestions for new articles or with comments about the articles we have printed (it is always gratifying to receive a large number of appreciative letters—many come, but we also get some letters suggesting that articles are technically too sophisticated or, in some instances, erroneous choices).

TABLE 3—CLASSIFICATION BY TECHNICAL DIFFICULTY

	Surveys	Creative Curmudgeon Articles	Others[a]	Total
Most Difficult	7	7	3	17
Some Difficulty	10	11	12	33
Not Difficult	2	3	24	29
Totals	19	21	39	79

[a] Review articles or books and general essays on all literature; excludes very short communications.

Table 4 shows the distribution of material by subject matter area. It extends only to the commissioned material, but I believe that a quick glance at the distribution probably mirrors the preferred reading area of most of the members of the Association. If not, do let me know.

The 1975 Issue

In the 1975 volume we anticipate having four survey articles. One will be a survey of the literature on research and development; another will be a survey of the studies on the impact of increased women's participation on the economy; a third will be a survey on recent work in government regulation in industry; and the fourth is most likely to be a survey on the empirical work done on economic migration.

We have under consideration quite a few interpretive essays. Most of these articles are commissioned; indeed, it is probably fair to state that all articles are sooner or later "commissioned" insofar as everything goes through the rather tedious and careful refereeing process with the outcome invariably requiring considerable refocusing and/or rewriting. I have used a great many members of the Association to assist me in this refereeing process. Perhaps no one is used more frequently than the members of my Board of Editors. Three are finishing their terms of service at the end of this calendar year. They are Abram Bergson (Harvard University), Harry T. Oshima (Rockefeller Foundation/University of the Philippines), and Arnold Zellner (University of Chicago). Each has been a gracious and generous

TABLE 4—CLASSIFICATION BY SUBJECT

Classification	Commissioned Survey	Creative Curmudgeon Essays	Other[a]	Total
01 General	1	—	4	5
02 Theory	5	6	4	15
03 Thought (Methodology)	—	2	17	19
04 Economic History	—	—	2	2
05 Comparative Systems	1	1	1	3
11–12 Growth and Development	1	5	—	6
21–22 Econometric, Statistical Theory, Statistics	2	—	1	3
31 Monetary Economics	—	2	1	3
32 Fiscal Economics	1	2	1	4
40–44 International Economics	2	2	7	11
50 Managerial Economics	—	—	—	0
60 Industrial Organization, Industrial Regulation	—	—	—	0
70 Agricultural Economics	1	—	—	1
80 Labor Economics	1	1	—	2
90 Applied Welfare Economics, Regional Economics	4	—	1	5
Totals	19	21	39	79

[a] Review articles on books, general essays on all literature.

colleague. I am personally most grateful to them and take this opportunity to note their service to the members of the Association. The foregoing is not to minimize the contributions made by the other members of the Board of Editors; fortunately for me, they serve at least another year. These include Marcus Alexis, (Northwestern University), Martin Bronfenbrenner (Duke University), Anne P. Carter (Brandeis University), Anne O. Krueger (University of Minnesota), Tibor Scitovsky (Stanford University), and William S. Vickrey (Columbia University). I am asking the Executive Committee to approve replacements for those who have left the Board, but until such action has been taken by the Executive Committee it is premature to list those names.

Finally, I want to thank the following individuals, plus four anonymous parties, for specific acts of assistance to me during this calendar year:

Irma Adelman
Gary Becker
Howard Bowen
James Buchanan
Jacob Cohen
William Comanor
Horace J. DePodwin
George Feiwel
Martin S. Feldstein
John Gurley
Andrew Hacker
James J. Heckman
Geoffrey C. Harcourt
Morton I. Kamien
Allen C. Kelley
Thomas Kelly
Kelvin Lancaster
Richard Leftwich
Asatoshi Maeshiro
Edwin J. Nell
Richard R. Nelson
William Nordhaus
Joseph Pechman
Trout Rader
Gaston Rimlinger
Warren S. Samuels
Reuben E. Slesinger
Brinley Thomas
Lester Thurow
Hirofumi Uzawa
Andrew Weintraub
Gordon C. Winston

MARK PERLMAN, *Managing Editor*

REPORT OF THE DIRECTOR
JOB OPENINGS FOR ECONOMISTS

Four years ago at these meetings, Allan M. Cartter projected a substantial oversupply of new Ph.D.'s during the 1970's and urged the Association to study measures that might minimize the painful readjustments necessary to overcome a near-permanent disequilibrium in the market (Cartter, p. 305–10). Generalized gloom and dissatisfaction prevailed about both the state of the market and its workings. As a result, in 1971 the Association adopted a resolution calling for a radical improvement of hiring practices in the profession and open listing of all employment opportunities.[1] Pursuant to that resolution, the Committee on Hiring Practices (F. Ray Marshall, Chairperson) was established. At last year's annual meeting, the Executive Committee approved the Committee's recommendation of a bimonthly publication for open listing of vacancies and resolved that "all members of the *AEA* have a professional obligation to list their openings with the Association."[2]

Two issues of *Job Openings for Economists* (*JOE*) have been published. The October issue contained 249 openings, and the December one contained 494. Six hundred thirty-one (631) different jobs have been listed, of which about 85 percent are academic. The fields of specialization most in demand were general economic theory, econometrics, urban economics,

Table 1—Specializations Most in Demand Ranked by Number of Citations

Specialization	Rank October	December	Combined
General Theory (020)	1	1	1
Econometrics (210)	2	2	2
Urban (930)	3	4	3
Business Finance (520)	6	5	4
Public Finance (320)	13	3	5

Table 2—Number and Percent of Job Citations by Field of Specialization

Classification Code	December Number	December Percent[a]	October Number	October Percent[a]	Total Number	Total Percent[a]
000—General Economics	121	21	66	21	187	21
100—Growth, Development	33	6	21	7	54	6
200—Economic Statistics	56	10	37	12	93	10
300—Monetary and Fiscal	63	11	30	9	93	10
400—International	31	6	25	8	56	6
500—Business Administration, Finance	77	13	37	12	114	13
600—Industrial Organization	41	7	26	8	67	7
700—Agriculture	24	4	10	3	34	4
800—Labor	49	8	23	7	72	8
900—Welfare, Urban, etc.	67	11	39	12	106	12
A00—Related Disciplines	0	0	0	0	0	0
B00—Administrative	14	2	4	1	18	2
Total	576	100	318	100	894	100

[a] Items may not add to 100 percent due to rounding.

---

[1] See "Minutes of the Annual Meeting," p. 474.

[2] See "Minutes of the Executive Committee Meeting," p. 455.

TABLE 3—PRINTING AND MAILING COSTS OF *JOE*

Item	October	December	Total
Address and Stamp	$ 17.00	$ 21.00	$ 38.00
Postage	341.40	796.00	1,137.40
Subtotal	$ 358.40	$ 817.00	$1,175.40
Delivery Service	1.95	1.95	3.90
Foreign Postage	105.47	147.66	253.13
Printing	483.05	1,039.45	1,522.50
Subtotal	948.87	2,006.06	2,954.93
Computer Time	51.13[a]	68.91	120.04
Total Costs	$1,000.00	$2,074.97	$3,074.97
Copies Printed	2,125	2,400	4,525
Cost Per Copy	$.47	$.86	$.68

[a] Estimated.

business finance, and public finance (see Tables 1 and 2). Each vacancy may be classified by more than one field of specialization. Hence, the number of field citations exceeds the number of job openings.

Participation by the "prestige" departments has been good. Of the top ten ranked schools in the Roose-Andersen report, eight have listed vacancies; of the next nine, six listed; of the next ten, seven; and of the last sixteen, eleven—for a total of 32 out of 45 (over 70 percent). Of the sixty-two departments in the Chairman's Group, forty have listed openings. If other departments had participated to this extent, marginal costs would have exceeded marginal revenue by a goodly sum. Printing and mailing costs alone (excluding salaries and overhead) were $.47 for the October issue and $.86 for the December one (see Table 3). Salary, overhead, general supplies, and other costs of *JOE* are not allocated from the general budget of the Association. However, the director estimates that total costs exceeded total revenue by about $12,000 during 1974. We began with 1,858 subscribers and now have 2,275.

No listing will be accepted for *JOE* containing material which violates federal civil rights laws. Institutions under censure by the *AAUP* may list vacancies, but the censureship will be noted. Listing a vacancy presumes that the position will not be filled until applicants have had a reasonable amount of time to respond (i.e., thirty days).

Listing is simple, fulfills a professional obligation to the Association, increases information available in the market, and may indicate good faith to other interested parties. All for no charge to the employer.

C. ELTON HINSHAW, *Director*

REFERENCES

A. M. Cartter, "Whither the Market for Academic Economists?," *Amer. Econ. Rev. Proc.*, May 1971, *61*, 305–10.

"Minutes of the Annual Meeting of the American Economic Association," Dec. 28, 1971, *Amer. Econ. Rev. Proc.*, May 1972, *62*, 470–74.

"Minutes of the Executive Committee Meeting of the American Economic Association," Dec. 27, 1973, *Amer. Econ. Rev. Proc.*, May 1974, *64*, 453–57.

REPORT OF THE COMMITTEE ON THE STATUS OF WOMEN IN THE ECONOMICS PROFESSION

During the past year the Committee has improved the operation of the market for economists, added to the information about the economics profession, and increased the effective supply of women economists.

The unique nature of this Committee should be noted. It is a "working committee," which this year has produced better information flow in the current job market, collected new, needed data, produced research results, planned and conducted a session at the *AEA* annual meeting, is in the process of producing a national research conference on occupational segregation, and has offered encouragement and support to women economists who have registered with us. The number of registered women economists is close to 1,400.

The need for this Committee, unfortunately, is far from being met. In recognition of this fact and the productiveness of the Committee, the Executive Committee of *AEA* in March 1974 changed the status of this Committee from that of an *ad hoc* one to a standing committee. On behalf of our Committee members, I wish to thank the Executive Committee for this expression of appreciation of our work and of faith in our continued performance.

I also want to thank the individual members of our Committee who have so diligently carried out the mandate of this Association. The Committee deliberately was made larger than some *AEA* committees and consists of ten members. It was designed carefully to have diversity by geographic location, age, sex, stage of professional development, and type of home-base institutions, including business, government, and academic.

On behalf of the total Association, I also wish to thank the Ford Foundation for its two grants to the *AEA* over the past three years which have enabled the rapid growth of this Committee's work. With this support, the Committee has been able to undertake important projects in line with its charge from the *AEA*. The nature of the experiment that the Ford Foundation financed was that this was to be part of the regular *AEA* structure; it was to pursue the goals set by *AEA* for the benefit of the total Association; and, if the experiment worked, the Committee and its activities would be continued by *AEA*. As we come to the end of the three-year start-up period, it is appropriate to recognize that this goal has been met and to recognize again the tremendous help that the Ford Foundation has been in making this quantum leap possible.

More specifically, the work of the Committee on the Status of Women in the Economics Profession (*CSWEP*) during this past year has included the following:

1. Four *CSWEP Newsletters* have been written during 1974 and sent to nearly 1,400 women economists. The newsletter is one way the Committee improves the working of the market by listing job openings for economists submitted by prospective employers. This activity fills a need most appreciated by agency and department heads, as our voluminous correspondence shows. The *CSWEP Newsletter* also lists requests for articles, conference and program plans and participants, grant and fellowship opportunities, and notices of regional activities of women economists. It is also a mechanism by which the Committee can ask the reader for help in various *AEA* projects. The newsletter clearly has helped to widen the informal network

and to support our women colleagues who often have felt isolated in our profession.

2. The roster of women economists developed and computerized by this Committee has been maintained this year and used as the basis for mailing the *CSWEP Newsletter*. It just recently has been updated by sending to all persons listed copies of the material they had previously supplied us and asking for the most recent information on their area of specialization, highest degree in economics, grade or rank, and address. The roster has been used to provide prospective employers with a list of women economists who meet specified criteria such as field, degree, and rank, although it is not known whether they are currently in the job market. (Whether they are or not may well depend on the opportunity.) The prospective employer may write to these economists, ask them to supply career data, and then follow up with interviews if desired. On an interim basis this service is currently being provided for a fee to help defray costs.

A special list of all women economists from the new *AEA Directory of Members* is in process of being run and will be used to add any new names to the roster list. Based on survey returns, the Committee feels that the number of women who have joined *AEA* has increased, directly or indirectly, because of Committee activity.

3. For the previous two years, the survey of academic economists was conducted by the Committee, and the survey results have been widely used. This year, on recommendation of this Committee, the collection of such data was transferred to the Nashville office of *AEA* and combined with the requests for other information from academic departments of economics into the Universal Academic Questionnaire. We participated actively in the development of the new survey and have continued as before to analyze the data related to the supply and demand for women economists. The preliminary results from this survey are presented below. Subsequent work will link these data to the previous data so as to begin to examine change, if any, over the three-year period of the life of this Committee.

4. A special survey of women and men economists in 1974–75 and their work histories is one of our major activities this year. The preliminary results from 710 women respondents and the first returns from a matched sample of male economists have been prepared. These results served as the basis for three research papers presented and discussed at the Annual Meeting session planned and conducted by the Committee. The male sample is still being developed, based on a random sample selected from the *Directory* to match the women respondents.

5. A national, working research conference on Occupational Segregation to be held in May 1975 is being sponsored and organized by this Committee. The conference is being held at the Center for Research on Women in Higher Education and the Professions at Wellesley College and is being financed by a grant from the Carnegie Foundation. Development and execution of this commitment will continue to involve much of our time. Publication plans are now being considered.

6. On-going efforts of the Committee this year have included the encouragement of the commissioning of a forthcoming survey article for the *Journal of Economic Literature* on the economic status of women, informal cooperation with similar women's groups of other professional associations, maintaining liaison with the Federation of Organizations for Professional Women, consultation with individual women economists or department chairpersons who have been concerned about particular cases of possible sex discrimination, answering requests, and urging program chairpersons of regional eco-

nomic associations and journal editors to appoint women to policy positions. Groups of women economists have been encouraged to meet together locally, and often informally, on an on-going basis. Efforts this year particularly have been made to increase the known supply of women economists in business and government.

Women in Academic Departments of Economics, 1974–75

Because no one knew how many women economists existed in academe or how many were currently being trained, this Committee for the last two years surveyed the academic departments of economics and agricultural economics. As noted above, this year similar data were collected under the direction of C. Elton Hinshaw by *AEA* in an expanded survey using the 1974–75 Universal Academic Questionnaire. The preliminary data that relate to women from the 1974–75 questionnaires returned to Nashville so far this year are summarized below. Only a few comparisons of overall totals from the three surveys are made; the data presented are not directly comparable because many departments did not provide data in all three years.[1] For example, last year's data related to 57 institutions in the *AEA*-designated Chairman's Group and this year's data to 47 in that group. Last year, in all, 414 departments with 4,262 full-time economics faculty members responded, and this year, so far, only 311 economics departments with 3,138 full-time faculty members have sent in their questionnaires.

In 1973–74, women comprised 16 percent of the students awarded degrees in economics at the bachelor's level, 14 percent at the M.A. level, but only 8 percent at the Ph.D. level. (See Table 1.) The proportions of women among senior majors in economics and candidates for M.A.s had not changed from the previous year, but the proportion of women receiving Ph.D.s in economics is considerably lower than the 12 percent reported the last two years for women Ph.D. graduate students.[2] It cannot be said that all four points of this difference represent a downturn in women Ph.D.s but this possible decrease bears watching.

The most common forms of financial aid to doctoral students in economics this year are "tuition and stipend with services required," and "tuition and stipend with no services required." Women Ph.D. students received such aid in proportion to their numbers (Table 2). The next most usual kind of financial aid is "stipend only with services required," and considerably less common, "tuition only with no service required." Disproportionately fewer women received these two forms of aid.

The 1974–75 employment of women who received Ph.D.s in economics in 1973–74 is similar to that of the men Ph.D.s except that more women are employed in academic institutions (68 percent of the women versus 56 percent of the men) and fewer women work in the federal government or are employed outside the United States (Table 3). Hinshaw suggests that the proportion of last year's Ph.D.s and *ABDs* finding employment in academic institutions this year is down considerably from the previous level.

The proportion of undergraduate women majoring in economics in Ph.D.-granting departments is much lower than in departments where the highest degree granted is an M.A. or a B.A. (This was also true last

---

[1] In the fall of 1973, questionnaires were sent to 1,409 departments of economics; in the fall of 1974, they were sent to 1,453 departments. Findings from last year's survey were reported by Carolyn Shaw Bell. Other preliminary findings from the 1974–75 survey have been reported by Hinshaw.

[2] The two previous surveys obtained data on students enrolled rather than degrees granted.

TABLE 1—DEGREES GRANTED IN ECONOMICS BY TYPE OF DEPARTMENT AND SEX, 1973–74

Departments Reporting by January 10, 1975, on 1974–75
Universal Academic Questionnaire

Degrees Granted in 1973–74	All Departments	Highest Degree Offered			
		Ph.D.			
		Chairman's Group	Other	M.A.	B.A.
Number of departments reporting	311	47	33	31	200
Ph.D., number	635	486	149	0	0
Percent women	8	8	9	—	—
M.A., number	1,348	691	418	239	0
Percent women	14	13	14	15	—
B.A., number	9,131	2,980	1,022	1,192	3,937
Percent women	16	12	10	23	18
Other degrees from economics departments, number	514	39	0	112	363
Percent women	14	13	—	17	13

TABLE 2—NUMBER OF FULL-TIME "ON-CAMPUS" GRADUATE STUDENTS REGISTERED FALL, 1974, AND TYPE OF FINANCIAL AID, BY TYPE OF DEPARTMENT AND BY SEX

Departments Reporting by January 10, 1975, on 1974–75 Universal Academic Questionnaire

Type of Department, Degree Sought, and Sex	Total	Receiving Financial Aid						No aid
		No Services Required			Services Required			
		Tuition only	Stipend only	Tuition and stipend	Tuition only	Stipend only	Tuition and stipend	
All departments:								
Ph.D. students, number	3,156	174	107	538	32	466	1,073	763
Female as percent of total	13	10	14	14	34	12	15	12
M.A. students, number	1,539	27	17	160	29	122	319	693
Female as percent of total	14	11	35	13	28	11	18	13
Chairman's Group:								
Ph.D. students, number	2,361	163	99	428	21	405	662	583
Female as percent of total	13	9	15	10	33	12	16	12
Ph.D., other departments:								
Ph.D. students, number	795	11	8	110	11	61	411	180
Female as percent of total	15	—	—	27	—	10	14	11
M.A. students, number	568	4	3	22	25	46	159	256
Female as percent of total	15	—	—	14	28	9	18	13
M.A. departments:								
M.A. students, number	186	12	13	29	4	1	38	78
Female as percent of total	17	17	46	10	—	—	24	8

Table 3—1974–75 Employment of 1973–74 Graduates in Economics by Level of Degree, Sex, and Type of Department

Departments Reporting by January 10, 1975, on 1974–75
Universal Academic Questionnaire

Type of Department and Kind of Employment	Ph.D.[a]		M.A.	
	Male	Female	Male	Female
All departments:				
Number	714	63	862	123
Percent	100	100	100	100
Percent employed as economist in United States:				
Educational institution	56	68	10	15
Business or industry	7	6	17	15
Federal Government	12	6	6	6
State/local government	3	5	6	6
Banking or finance	1	3	2	3
Percent employed outside United States:	12	1	14	7
Percent unemployed, seeking work	2	0	2	2
Percent not in labor force:				
Postdoctoral program	0	0	0	0
Entered or re-entered Ph.D. program	1	5	26	24
Not in school	3	5	2	7
Not known	2	0	15	9
Chairman's Group:				
Number	522	48	368	57
Percent	100	100	100	100
Percent employed as economist in United States:				
Educational institution	57	65	10	14
Business or industry	7	8	5	7
Federal government	11	8	6	4
State/local government	2	2	4	7
Banking or finance	1	2	2	4
Percent employed outside United States	11	2	14	12
Percent unemployed, seeking work	2	0	0	0
Percent not in labor force:				
Postdoctoral program	1	0	0	0
Entered Ph.D. program	1	6	35	28
Not in school	4	6	3	7
Not known	3	0	20	18

[a] Includes graduate students who have not completed their dissertations, if they entered the labor market seeking full-time employment as economists.

year.) This year the proportion is nearly 12 percent in Ph.D.-granting departments and nearly 20 percent in non-Ph.D.-granting departments. These overall data again suggest that the proportion of women undergraduate majors is positively correlated with the proportion of women faculty. The proportion of women among full-time faculty in tenure-track positions is 3 percent in Ph.D.-granting departments and 6 and 8 percent, respectively, in departments where the highest degree granted is an M.A. or a B.A. (Table 4).

Departments in the Chairman's Group have a low proportion of women among the full-time faculty in tenure-track positions at professor and associate professor levels (1 percent), but they have a much higher proportion of women at instructor and "other" levels (33 percent). The number of full-time faculty in non-tenure-track positions in the Chairman's Group is about the same as the number of part-time faculty members in those depart-

TABLE 4—NUMBER OF FACULTY BY RANK AND TYPE OF DEPARTMENT, 1974–75, BY SEX

Departments Reporting by January 10, 1975, on 1974–75
Universal Academic Questionnaire

Type of Appointment, Rank, and Sex	All Departments	Highest Degree Offered			
		Ph.D.			
		Chairman's Group	Other	M.A.	B.A.
Number of departments reporting	311	47	33	31	200
Full-time faculty, tenure-track:					
All ranks, male and female, number	3,138	1,187	679	353	919
Professors	1,225	609	270	114	232
Associate professors	816	269	208	99	240
Assistant professors	955	297	181	127	350
Instructors	129	9	18	13	89
Other	13	3	2	0	8
Female, percent of total	5	3	4	6	8
Professors	2	1	2	3	6
Associate professors	5	1	4	8	9
Assistant professors	7	7	5	6	9
Instructors	11	33	6	8	10
Other	8	33	0	0	0
Full-time faculty, non-tenure-track:					
All ranks, male and female, number	266	108	62	26	70
Professors	11	6	3	2	0
Associate professors	25	16	3	0	6
Assistant professors	100	42	18	8	32
Instructors	67	19	21	8	19
Other	63	25	17	8	13
Female, percent of total	14	15	21	4	9
Professors	0	0	0	0	0
Associate professors	4	0	33	0	0
Assistant professors	14	17	17	0	12
Instructors	9	11	10	0	11
Other	24	28	41	12	0
Part-time faculty:					
All ranks, male and female, number[a]	418	110	59	72	177
Professors	36	11	11	0	14
Associate professors	18	5	3	0	10
Assistant professors	49	11	4	9	25
Instructors	131	35	20	29	47
Other	184	48	21	34	81
Female, percent of total[b]	18	16	22	15	20
Professors	11	9	9	0	14
Associate professors	6	0	33	0	0
Assistant professors	22	45	25	11	16
Instructors	18	11	20	14	23
Other	21	17	29	18	22

[a] In all departments, fourteen of these positions are tenure-track, seven as professors. All seven of the professors are in departments that are in the Chairman's Group.

[b] In all departments, five of these positions held by women are tenure-track, one at each rank.

ments, each about 8 percent of the total faculty. Women hold 15 and 16 percent of these two types of positions. The non-Ph.D. departments, on the other hand, depend much more on part-time faculty (15 percent of the total faculty) and have a relatively high proportion of women in the part-time positions (nearly 20 percent versus about 2 percent in the full-time but non-tenure-track positions).

The continued existence of all-male economics faculties is of particular interest. Considering full-time faculty, 30 percent of the departments in the Chairman's Group have no women faculty (a similar proportion to that reported last year), 45 percent of the other Ph.D.-granting departments have no women faculty, and nearly 70 percent of the non-Ph.D.-granting departments are all male (Table 5). The proportion of departments with all-male part-time faculty (30 percent) exceeds the proportion with one or more women among the part-time faculty (18 percent).

Overall, the number of full-time economics faculty increased 2 percent from 1973–74 to 1974–75 (Table 6). This was an increase of eighty-six positions for the departments reporting, with thirteen of them being held by women (15 percent). Women represented 8 percent of the full-time faculty released (resignation, retirement, and nonrenewal of contract) and 9 percent of the new hires for full-time faculty positions. Of the overall net changes in full-time faculty, women represented a larger proportion of the increase in departments in the Chairman's Group than in other departments (23 versus 12 percent). This is

TABLE 5—DISTRIBUTION OF DEPARTMENTS OF ECONOMICS BY NUMBER OF WOMEN FACULTY MEMBERS, 1974–75

Departments Reporting by January 10, 1975, on 1974–75 Universal Academic Questionnaire

		Highest Degree Offered				
		Ph.D.				
Number of Women on Faculty	All Departments	Total	Chairman's Group	Other	M.A.	B.A.
Number of departments reporting	297	73	44	29	31	193
Percent distribution by number of women full-time faculty	100	100	100	100	100	100
All male	59	36	30	45	48	70
1 woman	29	44	48	38	36	23
2 women	7	8	11	3	16	5
3 or more women	4	12	11	14	0	2
Percent distribution of departments by number of women part-time faculty	100	100	100	100	100	100
No part-time faculty	52	41	41	41	42	57
1 or more part-time faculty	48	59	59	59	58	43
All male	30	31	27	38	29	30
1 woman	14	21	23	17	26	10
2 women	3	5	9	0	3	2
3 or more women	1	2	0	4	0	2
Total number of faculty						
Full-time	3,404	2,036	1,295	741	379	989
Percent women	6	4	4	5	6	8
Part-time	418	169	110	59	72	177
Percent women	18	18	16	22	14	20

TABLE 6—NET CHANGE IN FACULTY POSITIONS FROM END OF 1973–74 TO 1974–75, BY SEX, ALL DEPARTMENTS AND CHAIRMAN'S GROUP

Departments Reporting by January 10, 1975, on 1974–75 Universal Academic Questionnaire

Item	All Ranks	Professors	Associate Professors	Assistant Professors	Instructors	Others
All departments:						
Faculty released end of AY 1973–74:[a]						
Full-time, number	313	57	35	145	60	16
Women as percent of total	8	5	0	11	7	6
Part-time, number	90	5	3	12	34	36
Women as percent of total	8	0	33	0	3	14
New hires, faculty, AY 1974–75:						
Full-time, number	399	39	37	193	100	30
Women as percent of total	9	5	8	9	7	23
Part-time, number	158	5	3	24	52	74
Women as percent of total	16	0	0	13	8	26
Net change, 1973–74 and 1974–75:						
Full-time, number	+86	−18	+ 2	+48	+40	+14
Women, number	+13	− 1	+ 3	+ 2	+ 3	+ 6
Part-time, number	+68	0	0	+12	+18	+38
Women, number	+19	0	− 1	+ 3	+ 3	+14
Chairman's Group:						
Faculty released end of AY 1973–74:[a]						
Full-time, number	111	33	14	47	11	6
Women as percent of total	5	6	0	4	0	17
Part-time, number	29	2	0	0	13	14
Women as percent of total	7	0	0	0	8	7
New hires, faculty, AY 1974–75:						
Full-time, number	137	19	18	75	17	8
Women as percent of total	8	5	0	9	12	12
Part-time, number	38	0	1	4	18	15
Women as percent of total	11	—	0	25	6	13
Net change, 1973–74 and 1974–75:						
Full-time, number	+26	−14	+ 4	+28	+ 6	+ 2
Women, number	+ 6	− 1	0	+ 5	+ 2	0
Part-time, number	+ 9	− 2	+ 1	+ 4	+ 5	+ 1
Women, number	+ 2	0	0	+ 1	0	+ 1

[a] Resignation, retirement, and nonrenewal of contracts.

primarily due to the fact that five women represented 18 percent of the net increase at the assistant professor level in the Chairman's Group, but the number of women at that level in the other departments actually decreased. It should be noted that there was a net increase of three women at the associate professor level in the other departments.

Considering part-time faculty for all departments, the new hires of women exceeded the number released by nineteen and was primarily at the "other" rank. Most of this increase in part-time employment for women took place in departments that are not in the Chairman's Group.

The women employed as new hires for a given department this year tended to come less from other faculties, business, or government than did men, and to have come more from graduate students (Table 7). This is related to the fact discussed above that much of the increase in women hired has been for part-time jobs at lower faculty ranks. Among the women released

TABLE 7—PRIOR ACTIVITY OF NEW 1974–75 APPOINTMENTS AND PRESENT ACTIVITY OF "RELEASES" FOR 1973–74, BY TYPE OF DEPARTMENT AND SEX

Departments Reporting by January 10, 1975, on 1974–75 Universal Academic Questionnaire

Highest Degree Offered by Department and Activity of Faculty	New Hires in 1974–75ᵃ by Prior Year Activity		Those Released for 1974–75 by Present Activityᵃ	
	Male	Female	Male	Female
	(In Percent)		(In Percent)	
All departments	100	100	100	100
U.S. business and industry	5	0	13	9
Fed./state government in United States	3	0	15	4
Outside United States	1	0	6	0
Faculty at another school	34	31	40	52
Graduate student	54	59	8	13
Research leave	0	3	4	9
Retired	1	0	7	0
Other	2	8	7	13
Chairman's Group				
U.S. business and industry	2	0	7	0
Fed./state government in United States	3	0	20	0
Outside United States	3	0	7	0
Faculty at another school	32	16	43	83
Graduate student	55	62	4	0
Research leave	0	6	7	0
Retired	2	0	7	0
Other	3	16	5	17
Ph.D., other				
U.S. business and industry	8	0	17	29
Fed./state government in United States	1	0	14	14
Outside United States	1	0	2	0
Faculty at another school	41	45	41	29
Graduate student	47	55	9	14
Research leave	0	0	3	0
Retired	0	0	9	0
Other	1	0	5	14
M.A.				
U.S. business and industry	8	0	21	0
Fed./state government in United States	5	0	3	0
Outside United States	0	0	3	0
Faculty at another school	32	33	45	0
Graduate student	55	66	15	0
Research leave	0	0	0	0
Retired	0	0	3	0
Other	0	0	9	0
B.A.				
U.S. business and industry	6	0	14	0
Fed./state government in United States	3	0	16	0
Outside United States	0	0	8	0
Faculty at another school	31	33	33	50
Graduate student	57	56	9	20
Research leave	1	0	2	20
Retired	0	0	7	0
Other	2	11	10	10

---

[a] Includes full-time and part-time faculty.

last year by the reporting departments, none had retired (compared with 7 percent retirees among the released men). Relatively fewer women than men who were released went to business or industry or to positions outside the United States, but relatively more women went to other faculty positions. This pattern is similar to that reported above for types of employment for women with new Ph.D.s.

The economics departments reporting were asked this year to rank women full-time faculty by whether their salaries were above or below the departmental median for the particular rank and whether their length of service in that rank was above or below the median time at that rank for departmental faculty. Such estimates ignore how much the woman's salary is above or below the median. In these days when sex discrimination in pay is illegal, there undoubtedly is reluctance to report lower salaries for women. Nevertheless, the data reported reflect the situation as seen by department chairpersons or their reporting agents. For all departments, 36 percent of the women faculty were reported to have salaries more than $250 below the median (Table 8). When time in rank is also considered, 10 percent were below the median salary even though they were at or above the median time in rank. At the other end of the distribution, 20 percent of the women had salaries above the median with time in rank at or below the median time.

Women obtained 5 percent of all promotions (twelve of the 232), whereas they comprised 6 percent of all full-time faculty (Table 9). The promotions reported for women included one to professor and one to associate professor in the Chairman's Group (two out of sixty-nine, or nearly 3 percent), one to associate profes-

TABLE 8—RELATIVE SALARIES FOR RANK AND TIME IN RANK OF FEMALE FULL-TIME ECONOMISTS, 1974–75, BY TYPE OF DEPARTMENT

Departments Reporting by January 10, 1975, on 1974–75 Universal Academic Questionnaire

Highest Degree Offered by Department and Relative Salary for Rank	All Women	Time in Rank		
		Above Median	At Median	Below Median
		(In Percent)		
All departments	100	24	33	43
Salary above median	31	19	6	8
Within $250 of median	33	3	21	6
Salary below median	36	2	5	29
Ph.D., Chairman's Group	100	22	34	44
Salary above median	26	18	4	4
Within $250 of median	28	2	20	6
Salary below median	46	2	10	34
Ph.D., other	100	12	19	69
Salary above median	23	4	4	11
Within $250 of median	36	0	12	23
Salary below median	41	8	4	34
M.A.	100	39	33	28
Salary above median	50	39	6	6
Within $250 of median	22	0	22	0
Salary below median	28	0	6	22
B.A.	100	26	36	38
Salary above median	39	20	9	9
Within $250 of median	34	5	26	3
Salary below median	27	0	1	26

TABLE 9—PROMOTIONS AND TENURE DECISIONS FROM 1973–74 TO 1974–75
BY TYPE OF DEPARTMENT AND SEX

Departments Reporting by January 10, 1975, on 1974–75
Universal Academic Questionnaire

Highest Degree Offered by Department and Rank	Promotions to Rank		Given Tenure at Rank	
	Total Number	Female as Percent of Total	Total Number	Female as Percent of Total
All departments	232	5	148	3
Professor	77	8	22	0
Associate professor	118	3	86	3
Assistant professor	36	3	35	3
Ph.D., Chairman's Group:				
Professor	29	3	8	0
Associate professor	40	2	30	0
Assistant professor	4	0	3	0
Ph.D., other:				
Professor	14	0	6	0
Associate professor	25	4	18	0
Assistant professor	5	0	4	0
M.A.:				
Professor	10	20	2	0
Associate professor	12	0	9	0
Assistant professor	2	0	8	0
B.A.:				
Professor	24	13	6	0
Associate professor	41	5	29	10
Assistant professor	25	4	20	5

sor in the other Ph.D.-granting departments, and six promotions in departments where a B.A. is the highest degree granted.

Effective affirmative action at the department level is probably more a matter of action by men of good will and sensitivity who try to act in good faith to open opportunities to women economists and to change traditional attitudes so as to offer better educational opportunities for women than it is a matter of coerced action. Undeniably, however, the threat of coercion has increased men's awareness. In some educational institutions, Affirmative Action (*AA*) offices at the central administrative level have been effective in increasing awareness of legal and moral responsibilities and in facilitating implementation of *AA* plans. At the department level, perception of the effectiveness of *AA* offices may well be a combination of the particular experiences of the economics department and a projection of the values of the department chairman or his agent who answers the questionnaire. In all, 71 percent of the departments were in institutions with *AA* offices (Table 10). Of these, 80 percent had been in some type of contact with the office. Of these, 55 percent found the *AA* office useful, 35 percent found it "not at all useful," and 10 percent found it a "damned nuisance." Among the Chairman's Group, all departments were in institutions with *AA* offices, and 91 percent had been in contact. Of these 35 percent found the *AA* office to be useful, 42 percent "not at all useful," and 23 percent, a "damned nuisance." Resentment of *AA* appears to be greater among department chairpersons in the Chairman's Group than in other departments of economics.

It is not often that authors of an article have a chance to quantify the proportion

TABLE 10—AFFIRMATIVE ACTION AND ECONOMICS DEPARTMENTS, FALL 1974

Departments Reporting by January 10, 1975, on 1974–75
Universal Academic Questionnaire

Item	All Departments	Highest Degree Offered			
		Ph.D.			
		Chairman's Group	Other	MA	BA
Number of departments reporting	301	47	32	31	191
Departments in institutions with Affirmative Action (*AA*) Office:					
Number	213	47	25	30	111
Percent	71	100	78	97	58
Percent of departments in institution with *AA* office who have been in contact with *AA* office	80	91	88	83	72
Percent in contact finding services of *AA* office—					
Very useful	11	12	22	10	8
Somewhat useful	44	23	39	59	68
Not at all useful	35	42	35	17	37
Damned nuisance	10	23	4	16	5
Percent of department heads who have read December 1973 *Amer. Econ. Rev.* article, "Combatting Role Prejudice and Sex Discrimination"	41	49	41	67	38
Of these, percent who have used any of the suggestions of the article	68	64	67	50	75

of department chairpersons who have read a given article. In the December 1973 issue of the *American Economic Review*, Kenneth E. Boulding and Barbara B. Reagan on behalf of this Committee wrote an article, "Combatting Role Prejudice and Sex Discrimination." (Informally among Committee members this article was referred to as guidelines for good guys.) Over 40 percent of all department chairpersons and nearly 50 percent of the Chairman's Group reported that they had read this article, and 68 and 64 percent, respectively, said that they had used suggestions of the article.

BARBARA B. REAGAN, *Chairperson*

SUMMER PROGRAM IN ECONOMICS FOR MINORITY STUDENTS

The Summer Program in Economics for Minority Students was originally conceived in 1969 in discussions between the Caucus of Black Economists and the Executive Committee of the American Economic Association. Members of the Caucus were distressed about the small numbers of minorities in the economics profession and concluded that this shortage could only be remedied by an intensive program of solicitation and education of young minority men and women who had not completed their undergraduate education and who showed signs of promise as evidenced by their work in economics and related subjects. The Caucus and the American Economic Association's Executive Committee agreed that a committee should be formed to investigate the matter in depth. Such a committee was formed and began functioning in January in 1970. The Committee conducted an in-depth survey of minority holders of advanced degrees in economics and of minority students studying economics on the undergraduate level in the major black institutions; it also surveyed the curriculum of these institutions. After the survey, it was even more clear that if minority students were to have any prospects of being admitted into and performing at a high level in the most competitive programs in economics, many of them would require additional work in economics theory and quantitative methods.

In the Summer of 1971 members of the American Economic Association Committee on Minorities and Members of the Caucus of Black Economists met in Boston under the chairmanship of Robert Dorfman of Harvard University to discuss plans for establishing one or two summer institute programs of intensive work in economic theory and quantitative methods for minority students. Because of difficulties in making adequate financial arrangements and the unavailability of a suitable host institution, these sessions did not culminate in a summer program. A similar effort in 1972 was also aborted. Thus, three opportunities (1970, 1971, 1972) for minority students to obtain suitable backgrounds to enter the mainstream of economics education at the graduate level eluded the efforts of the American Economic Association's committee. A similar fate was to befall a proposed institute in the summer of 1973.

In 1974, Marcus Alexis, a charter member of the Caucus of Black Economists and a member of the American Economic Association Committee on Minorities who teaches at Northwestern University, took it upon himself to move the project forward, fearing that a fifth year might pass without a summer program for minority students.

In January 1974 he inquired of Richard Holton, Dean of the Graduate School of Business at the University of California, Berkeley, if the Berkeley campus would be willing to host a Summer Program for Minority Students, sponsored by the American Economic Association and funded by outside agencies. Receiving an affirmative reply, he communicated the information to Robert Dorfman, chairman of the committee. Dorfman was enthusiastic and at the March meeting of the Executive Committee of the American Economic Association formal approval was given to seek financing for the Summer Program and to solicit applications from minority students.

The Alfred P. Sloan Foundation con-

TABLE 1—BUDGET ALLOCATION FOR SUMMER PROGRAM IN ECONOMICS FOR MINORITY STUDENTS, 1974

Students	
Stipend—22 @ $200	$ 4,400.00
Room and Board—22 @ $500	11,000.00
Transportation	8,204.79
Faculty	
Salaries (includes teaching assistants)	18,403.08
Transportation	2,589.20
Teaching Materials	6,024.46
Administration and Facilities	7,372.83
	$57,994.36

tributed $20,000 towards a proposed Summer Program. Inquiries made at the Ford Foundation produced encouraging responses. The Ford Foundation later made a grant for the Summer Program of $25,000. Through the efforts of several members of the Executive Committee of the American Economic Association, additional funds were received from other interested parties.

Because of the late start (March 1974), it was necessary to launch an intensive program campaign immediately to interest students, to assemble a staff, and to make administrative arrangements with the Berkeley administration. After three months of frenetic activity, the Summer Program for Minority Students became a reality.

Recruitment

On the basis of letters of recommendation and student transcripts, thirty students were offered positions in the Summer Program even though funds had been secured in amounts to finance only twenty-five. In the end, twenty-five students did accept the invitation to be members of the Summer Program. But by the time the program began on June 20, three of the students had to withdraw. Twenty of the students were black, one was of Latin extraction, and one was a mixture of black and Latin. The students came from a wide variety of schools and were well represented geographically. Twelve of the twenty-two students came from predominantly black colleges. Of these, five were from the Atlanta University complex, two from Benedict College (South Carolina), and one each from Fisk University (Tennessee), Talladega College (Alabama), Tougalloo College (Mississippi), Bishop College (Texas), and Savannah State (Georgia). The remaining ten students came from Cornell University (New York), Oklahoma State University, City College of New York, University of Massachusetts (Boston), University of Pennsylvania, LaSalle College (Pennsylvania), Oakland University (Michigan), Northeastern Illinois University, Mills College (California), and Hayward State College (California).

No attempt was made to predict the performance of the students in the program. With one exception, all of the students had cumulative grade point averages of 3.0 or better. Each student was strongly recommended by at least one faculty member from his or her institution. The small number of students did not permit any formal experiment to ascertain whether students from one type of college would perform best. The final results showed, however, that students from predominantly black institutions performed as well as those from majority white schools. This limited evidence suggests a large reservoir of untapped talent in these schools.

Curriculum

The curriculum in the Summer Program consisted of microeconomics, macroeconomics, quantitative methods, and expository writing. In microeconomics students were expected to master material at the level of Edwin Mansfield's *Microeconomics* text, and William J. Baumol's *Economic Theory and Operations Analysis.* In macroeconomics the primary text was Thomas F. Dernburg and Duncan M. McDougall, *Macroeconomics.* The text in quantitative methods was Alpha Chiang's *Fundamental Methods of Mathematical Economics.* The instruction was evenly divided between microeconomics, macroeconomics, and quantitative methods. Students spent a minimum of fifteen hours in class each week. There were five hours of lecture per week in each of the three subjects and often supplementary sessions. In addition there were ten hours of laboratory work in which students could receive assistance in troublesome assignments from three teaching assistants hired for the program. Microeconomics was taught by Marcus Alexis of Northwestern; macro by Terry Turner of the University of California, Berkeley; and quantitative methods by Glenn Loury, a third-year graduate student at Massachusetts Institute of Technology. The three teaching assistants provided a total of sixty hours per week. They were carefully selected by Turner and they proved to be outstanding. Each of them was a fourth-year graduate student in applied economics at the University of California, Berkeley.

In addition to the formal lectures and laboratory sessions, there were three simulated economic games developed in the Behavioral Science Laboratory of the Center for Research in Management Science at the University of California, Berkeley. Leonard Doyle of the University of California, Berkeley faculty assisted in the briefing of the students and the write-ups explaining the purpose of the economic games. The simulation was an immense success, both because the students enjoyed them and because they did so well in the exercises. Another factor in the success in the simulations was that they gave the students an opportunity to apply the concepts and tools they had acquired in the more formal classroom sessions and in homework. In future programs, more of an attempt should be made to provide students with opportunities to apply their knowledge to problems of policy and solutions to economic problems faced by professional economists.

Seminars and Counseling

The Summer Program was fortunate also in having a very successful set of seminar speakers. Charles E. Wilson, Vice-Chancellor of the University of California, Los Angeles, was the first visitor and spoke to the students of opportunities in economics for minority students and of the use of economics in solving administrative problems in complex organizations. Durand Bell of the University of California, Irvine, spoke on the "Minimum Wage Reconsidered." Donald J. Harris, University of Wisconsin, Madison, spoke on "A Critique of Neoclassical Growth Theory." Richard Sutch, professor of economics at the University of California, Berkeley, gave an in-depth analysis of *Time on the Cross*, a volume which the students had read in preparation for the seminar. Finally, Robert Aaron Gordon of the economics department at the University of California, Berkeley, discussed prospects for careers in economics and presided at the closing session of the Summer Program.

In addition to the formal classwork, seminars, laboratory sessions, simulations, and homework assignments, a very important aspect of the Summer Program was the counseling that students received. Most of the students knew little about the offerings of the various graduate programs in economics and of the relevant standings of various departments. Very early in the program an effort was made to prepare students for the demands that would be placed upon them by the better economics departments. The schedule of the Summer Program was intentionally made very intense in an attempt to simulate conditions in graduate schools. This afforded an opportunity for those students who would be discouraged at such a work load to reflect on whether they were willing to make the sacrifice in time and effort required for success in the more competitive economics departments, and for the students and the faculty to evaluate their potential as future graduate students in economics. The counseling consisted of formal and informal discussions of graduate programs available and the emphasis in these programs. The counseling included such routine matters as how to solicit information from graduate schools, financial assistance, completion of applications, and candid appraisals of students' chances for admission and success in schools in which they showed interest.

Postprogram activity has included contacting schools on behalf of the students, writing letters of recommendation, and advising students on the appropriate courses they should take in their fourth year of college. The latter was extremely important because most of the students were entering their senior year, many with the need of additional course work in mathematics, statistics, and courses in economic theory and analysis. Letters from the students since the program indicate that this has been well received.

MARCUS ALEXIS

REPORT OF COMMITTEE ON ECONOMIC EDUCATION

In 1974, this Committee continued to develop the program laid out in its 1972 agenda for improving the teaching of economics (published in the *American Economic Review* in May 1973). We wish to call the attention of the profession especially to three developments.

(1) The project outlined in the 1972 agenda to help university economics departments provide improved training in teaching for new Ph.D.s has been broadened and extended to five universities. The goal of the program is to help potential teachers on both techniques of teaching and the subject matter of basic economic courses which they will probably be asked to teach; greater details are provided in the May 1973 issue of the *Review*.

During 1973–74, the Universities of Minnesota and Wisconsin used the new program with substantial success. On the basis of this experience, further improvements have been made in both the written and video tape materials designed to help potential teachers, and they are being used in four additional schools in 1974–75—Florida State, Indiana, Nebraska, and Purdue Universities. Current plans for the program involve extending the experiment to at least five more schools over the two years ahead.

The Committee hopes that these experiences, teaching materials, and information on teaching techniques for new teachers may be useful to a wide variety of universities training future teachers. Further information can be obtained by writing to Dr. Arthur Welsh, Joint Council on Economic Education, 1212 Avenue of the Americas, New York, New York 10036.

(2) The first of the new *Readings* series, designed to help teachers at smaller institutions keep up-to-date on central areas of economic theory and policy, has appeared under the title *Recent Advances in Economics*. The series, to be published by Richard Irwin, Inc., is under the editorship of Rendigs Fels and John Siegfried of Vanderbilt University and the general policy guidance of this Committee. The readings have been selected especially to help teachers of economics at community and junior colleges and at other less-known colleges and universities where faculty members often find it difficult to keep up with new theory and policy developments and to do an adequate job of teaching a variety of courses. Each annual volume will include a series of articles, reprinted from economic journals and other sources, aimed at pulling together at one place materials that may be especially useful.

(3) At the 1974 meetings of the Association in San Francisco, this Committee sponsored an exhibit on innovations in the teaching of undergraduate economics. The Committee invited teachers who have been experimenting with new techniques and technological developments in the teaching of economics to exhibit their experiments in a "hands-on" mode so that economists visiting the convention could see for themselves how the new developments work. The primary goal was to make information on these new developments available as widely as possible throughout the profession.

The exhibits, together with the individuals from whom further information can be obtained on each, are indicated below:

1. *Macro Simulation Models for Use with Time-Sharing Computers.* A simple model of the economy is put on the com-

puter, and students are able to change the tax rate, government expenditures, and the money stock in attempts to stabilize a moderately complex economy. Different welfare functions are given to students to be maximized. (Richard Attiyeh, University of California, San Diego, and Keith Lumsden, Stanford University.)

2. *Teaching Information Processing System* (*TIPS*). An extensive review of basic economics has been programmed on a computer so that students can review for themselves at their own pace how well they are learning the material contained in traditional economics texts. In effect, the computer substitutes for a combination workbook and teaching assistant. (Allen Kelley, Duke University.)

3. *Study Management System* (*PLATO*). This system is similar to *TIPS* in that it permits students to study at their own pace the subject matter of a standard elementary course. It includes individual study and review material. Students, using the computer, answer a series of examination questions and are either told to go on if they perform satisfactorily or helped to review individually if they need further assistance. (Don Paden and Bruce Delgaard, University of Illinois.)

4. *Computer-Aided Instruction Modules in Intermediate Macroeconomics.* This is a series of modules in teaching intermediate macroeconomics, stressing problems and situations involving computations—for example, accelerator-multiplier models, monetary models, and the like. They are designed for teachers who want to save class time for other activities. (James Weatherby and James Peach, University of Texas at Austin.)

5. *Open University Economics Courses.* In the United Kingdom, open university courses have been offered in economics as well as other fields. TV instruction and cassettes are used for audio-visual teaching, plus newly developed reading materials. While the materials are designed for United Kingdom students, a number of the modules can be readily adapted for American use at the elementary level. (Leslie Wagner and Anthony Mellor, The Open University, Walton Hall, Walton, Bletchley, Buckinghamshire, London, England.)

6. *Teacher Training Program for New Ph.D.s.* Materials are now available, including extensive written suggestions and video tape materials, on the new program to aid Ph.D.s preparing for college and university teaching posts, described in the first paragraph above. Most of these materials have been developed by the group responsible for the program; some are specifically drawn from course experiments at the University of Minnesota, where video tapes have been used with especially good effect. (For basic materials, write Dr. Arthur Welsh, Joint Council on Economic Education, 1212 Avenue of the Americas, New York, New York 10036; additional information can be obtained from William Becker and Michael Salemi, University of Minnesota, and Phillip Saunders, University of Indiana.)

7. *Interdisciplinary Introductions to Economics.* Combined programmed learning books and video tape "grabbers" to get students involved in material being taught at different stages of an elementary course have been developed. These video tape introductions may be useful at other universities. (John Kuhlman, University of Missouri at Columbia.)

8. *Video Cassette Modules for Basic Economics.* Video tape modules are being developed for students in evening classes, especially adult students attending school on a part-time basis. Units are at an elementary level. (Douglas Poe, Illinois State University.)

This experimental exhibit was made possible by a generous grant from the Calvin Kazanjian Economics Foundation, which has for many years provided prizes for outstanding economics teachers at grade school, high school, and college levels, and which is prepared to provide copies of the winning teaching projects at different levels to those who write for them. The address is 251 Danbury Road, P.O. Box 452, Wilton, Connecticut 06897. The Committee also is especially indebted to Keith Lumsden of Stanford who undertook the responsibility of arranging the exhibit and seeing to its implementation.

G. L. BACH, *Chairperson*

REPORT OF THE ECONOMICS INSTITUTE POLICY AND ADVISORY BOARD

Each year since its inception in 1958 the Economics Institute has provided an intensive summer program of academic training and orientation for international students who are preparing to begin graduate study in the United States in economics and related disciplines, including business and agricultural economics. Continuing under the dedicated directorship of Wyn Owen, the Institute enrolled a record student body of 222, representing fifty-three foreign countries, in 1974.

Each summer the Institute offers intensive academic course work in economic theory, statistical methods, mathematics, business, and English, supplemented by special lectures, seminars, and field trips. By virtue of its increased size, the Institute was able to respond to more diverse student needs in recent years by providing an eleven and a half weeks extended session, a nine and a half weeks main session, and five and a half weeks short session. In 1974, the Institute also sponsored three associated projects: a conference in which doctoral candidate foreign students could discuss their research with each other and with current Institute students; the third edition of *Guide to Graduate Study in Economics and Agricultural Economics in U.S.A. and Canada;* and a research project on academic performance of Institute alumni.

The Institute was initiated with the aid of a grant from the Ford Foundation. Ford continues to support the program by providing a revolving guaranty fund and with a limited amount of fellowship aid. The program is, however, essentially self-supporting from tuition and fee charges substantially paid by a total of forty-eight sponsoring agencies, twenty-eight of them foreign.

The Institute now has a total of 1,654 alumni, many of whom are in important academic, government, and business posts throughout the world. The success of its alumni in their academic programs and their subsequent careers attests to the value of the Economics Institute. It is anticipated that the Institute will have an important role to play in international education for the foreseeable future.

The current members of the Board are Irma Adelman, Walter P. Falcon, Edwin S. Mills, Daniel Schydlowsky, Arnold Harberger, and Richard Holton.

EDWIN S. MILLS, *Chairperson*

REPORT OF REPRESENTATIVE TO THE INTERNATIONAL ECONOMIC ASSOCIATION

During the calendar year 1974, the International Economic Association held a conference in Italy, a world congress in Hungary, a workshop in Turkey, and an East-West meeting in Switzerland.

The conference in Turin, Italy, April 1 to 6, was on "The Economics of Public Services." It was prepared by a program committee chaired by Martin Feldstein (United States) with Anthony Atkinson (United Kingdom), Francesco Forte (Italy), Eugeny Gromov (*USSR*), Serge-Christophe Kolm (France), and Jean Paelinck (Netherlands) as members. Besides Feldstein, nine Americans participated: Henry Aaron, David Bradford, Harold Hochman, Robert Inman, Alvin Klevorick, Fritz Machlup, Mark Pauly, Joseph Stiglitz, and William Vickrey.

The World Congress in Budapest, Hungary, was held from August 19 to 24 on "Economic Integration: Worldwide, Regional, Sectoral." The Program Committee, chaired by Fritz Machlup, included: Bela Balassa (United States), Raymond Barre (France), Oleg Bogomolov (*USSR*), Béla Csikós-Nagy (Hungary), Robert Gardiner (Ghana), Herbert Giersch (Federal Republic of Germany), Harry Johnson (United Kingdom), Ryutaro Komiya (Japan), Erik Lundberg (Sweden), Joszef Pajestka (Poland), and Raul Prebisch (Argentina). The program provided for papers and reports in plenary sessions and for papers and general discussion in ten separate working groups. Approximately 1,800 economists from 77 countries attended and over 500 participated in the discussions. The program contained the names of 82 economists (from 42 different countries), 53 as authors or coauthors of main papers, discussion papers, and opening statements, and 29 as chairmen or vice-chairmen of sessions and working groups. The largest delegation came from the *USSR* with 280 members. The United States was represented by 78 participants; of these the following six presented papers or discussion papers: Bela Balassa, Richard Cooper, Gottfried Haberler, Peter Kenen, Fritz Machlup, and Robert Mundell. This was the Fourth World Congress of the International Economic Association; the previous congresses had been held in Rome (1956), Vienna (1962), and Montreal (1968). Future congresses will be held every third year, the next one, 1977, in Tokyo.

The workshop at the Bosphorus University in Istanbul, Turkey, was held September 2 to 6 on "Population and Development." The Program Committee consisted of Paul Demeny as chairman and Emre Gönensay (Turkey) and Frederic C. Shorter as members. It was attended by twenty-seven persons, of whom fourteen came from Asia, five from Africa, four from Europe (three of them from socialist countries), and one from South America. From the United States, Paul Demeny, Frederic Shorter, and Jeffrey McNicoll were present.

The East-West Conference was held from August 29 to September 3 in Montreaux, Switzerland, on "Recent Economic Developments in Socialist and Non-Socialist Countries." It was arranged by John T. Dunlop. Of the nineteen participants, eight came from Eastern and eleven from Western countries. From the United States, Abram Bergson, James Duesenberry, John Dunlop, Walter Galenson, and Fritz Machlup attended.

The plans for 1975 include a conference

on "The Microeconomic Foundations of Macroeconomic Theories," to be held near Barcelona, Spain, in April, planned by a Program Committee chaired by Geoffrey Harcourt (Australia), assisted by Sir John Hicks and four others; a conference on "The Organization and Retrieval of Economic Knowledge" in Kiel, Germany, in July, arranged by a committee with Herbert Giersch (Federal Republic of Germany) and Mark Perlman as cochairmen; and a conference on "Inflation Theory and Anti-Inflation Policy" in Saltsjöbaden, Sweden, at the end of August, programmed by a committee chaired by Erik Lundberg (Sweden).

Five volumes of proceedings were published during 1974: *Agricultural Policy in Developing Countries*, edited by Nurul Islam; *The Economics of Health and Medical Care*, edited by Mark Perlman; *The Economic Development of Bangladesh*, edited by E. A. G. Robinson and Keith Griffin; *The Management of Water Quality and the Environment*, edited by Jerome Rothenberg and Ian G. Heggie; and *Industrial Management: East and West*, edited by Aubrey Silberston and Francis Seton. The first four volumes were published in London by Macmillan and in New York by the St. Martin's Press. The fifth volume was published in New York by Praeger Publishers. Several other volumes are at the printers or in preparation.

The officers of the Association for the next three years are Edmond Malinvaud (France), President; Shigeto Tsuru (Japan), Vice-President; and Herbert Giersch (Federal Republic of Germany), Treasurer.

FRITZ MACHLUP

REPORT OF THE REPRESENTATIVE TO THE NATIONAL BUREAU OF ECONOMIC RESEARCH

During 1974, research operations began at the new office of the National Bureau of Economic Research-West in Palo Alto, California, under the codirectorship of Victor Fuchs and Sherman Maisel. By the end of the year, some thirty staff members and associates were at the new facility, many having transferred from the New York office. Two special conferences were held on the West Coast, dealing with Regional Stock Exchanges in a Central Market System and International Trade, Finance, and Development of Pacific Basin Countries. In addition to branch offices already established in Cambridge, Mass., and New Haven, Conn., an office was opened in Washington, D.C. to accommodate ten staff members and associates conducting research and conference programs in that area.

The research program during the year included more than one hundred individual projects. Reports on their status appear in the National Bureau's 54th Annual Report, September 1974, which may be obtained from the Bureau on request (261 Madison Ave., New York, N.Y., 10016).

The National Bureau was involved in three major conferences in 1974, the proceedings of which will be published. Future conferences are to be held on the Economics of Urban Spatial Structure, Gregory K. Ingram, Chairman; Economic Analysis of Political Behavior, Melvin W. Reder, Chairman; Causes and Economic Effects of Population Change in Less-Developed Countries, Richard A. Easterlin, Chairman; and Low-Income Labor Markets, Sherwin Rosen, Chairman. In addition, there were eight workshops centered about the use of the computer, including three in Latin America, and eleven working groups operated through the Conference on Econometrics and Mathematical Economics.

At the request of the National Science Foundation, the National Bureau undertook a new activity to assist in the development of a United States-*USSR* Program on Scientific and Technical Cooperation. The Bureau is helping plan and coordinate the fields of econometric modeling, economic programming, and computer analysis applied to the economics and management of large systems. Joint meetings were held in each country in 1974 and further meetings and exchange visits by specialists are planned for 1975.

A new quarterly journal, ***Explorations in Economic Research***, was inaugurated with the first two numbers issued in September and December, 1974. This journal is designed to present staff research reports of the type that were formerly published as occasional or technical papers. In addition to major articles, the journal will regularly present a summary of the American Statistical Association-National Bureau of Economic Research Business Outlook Survey.

At the Annual Meeting on September 30, 1974, J. Wilson Newman was elected Chairman of the Board succeeding Walter W. Heller, Moses Abramovitz was elected

Vice-Chairman, and John R. Meyer was reelected President. Eli Shapiro had been elected a Director-at-Large, filling the vacancy created by the death of Eli Goldston. New Directors-at-Large elected at the Annual Meeting were Andrew F. Brimmer succeeding J. Irwin Miller, Frank L. Fernbach succeeding Joseph A. Beirne, and Rudolph A. Oswald, succeeding Boris Shiskin. Paul W. McCracken was elected Director Representative of the American Statistical Association. Among Directors by University appointment, J. C. La Force, University of California, Los Angeles, was elected to succeed Thomas A. Wilson, University of Toronto. At the end of 1974, Hal B. Lary retired from his position as Vice President-Research.

WILLARD L. THORP

REPORT OF THE CENSUS ADVISORY COMMITTEE

The Census Advisory Committee met twice during 1974 (April 11–12, and October 17–18). On each occasion, Vincent P. Barabba, Director, and other members of the Census Bureau staff presented a variety of reports concerning long- and short-range Census projects and problems. The Committee provided the Census Bureau with a number of specific recommendations with respect to individual projects, data collection techniques, and program development. Copies of specific recommendations can be obtained from the Director or the Committee Chairperson. In addition, at the October meeting two subcommittees were established as follows: Review of Overall Plans for Economic Censuses—Anne P. Carter, Chairperson; and Measurement of Inventories—Andrew F. Brimmer, Chairperson.

Additional committee members will be appointed to these subcommittees at a later date, and subcommittee chairpersons will be in contact with other members of the Association with respect to problems and suggestions. Voluntary responses for Association members are hereby solicited and may be sent to each subcommittee chairperson directly. Comments or suggestions of Association members with respect to the Census Bureau in general or with respect to other specific areas of study may be sent to the committee chairperson at any time.

The following Committee members complete their terms of service at this time: Guy E. Noyes, Jacob Mincer, Patricia Shontz, Phyllis A. Wallace, and the current chairperson. The Committee will be chaired for 1975 by Richard Ruggles.

LEE E. PRESTON, *Chairperson*

◆ ◆ ◆

MAKE PLANS NOW

to attend the

1975 Meetings - Allied Social Science Associations

in

Dallas, Texas

Sunday, Monday, Tuesday,

December 28-30

Employment Center opens Saturday, December 27

This is a change from the October dates
previously announced.

Please note the 1976 ASSA meetings will be held
in Atlantic City, September 16-18.

◆ ◆ ◆

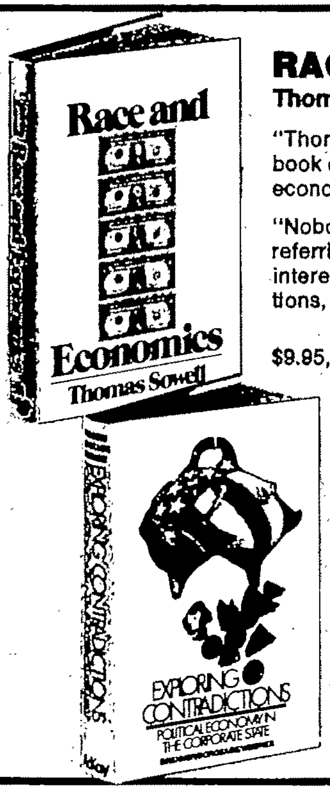
ECONOMIC INQUIRY
Journal of the Western Economic Association

Volume XII, Number 4 December 1974

A Neo-Classical Theory of Keynesian Unemployment . . . . *Donald F. Gordon*
Information Costs and Markets . . . . *Chung H. Lee*
Insurance Theory: Reserves Versus Mutuality . . . . *John M. Marshall*
On the Structure of Trade . . . . *Mordecai Kurz and Robert Wilson*
Money and the Production Function . . . . *Stanley Fischer*
A Very Unsubtle Version of Arrow's Impossibility Theorem. . . . . *Allan M. Feldman*
The Effect of the Bracero on Agricultural Production in California . . . . *Donald E. Wise*
Utility Functions with Jump Discontinuities: Some Evidence and Implications from Peasant Agriculture . . . . *Robert Tempest Masson*
The Neoclassical Dichotomization of Economic Theory . . . . *Will E. Mason*
A Reformulation of the Law of Demand . . . . *Ron Heiner*

Edited by Robert W. Clower
Department of Economics, University of California, Los Angeles

Correspondence relating to membership, subscriptions, and other business matters should be addressed to the Western Economic Association, Executive Office, Department of Economics, California State University, Long Beach, California 90840. Membership dues, including *Economic Inquiry*: one year, $12; three years, $33. Student membership, $7. Subscriptions for libraries, institutions, and business firms, $20. Add $2 for memberships or subscriptions outside the United States, Canada and Mexico.



RACE AND ECONOMICS
Thomas Sowell, *University of California, Los Angeles*

"Thomas Sowell's *Race and Economics* is, quite simply, the very best book ever published on the interrelationships of race, ethnicity, and economic achievement."—Irving Kristol, *New York University*

"Nobody can write or talk seriously about ethnicity or race without referring to this book.... Sowell is distinguished by the breadth of his interests, the discipline of his analysis, and that rarest of combinations, passion and fairmindedness."
—Edward Banfield, *University of Pennsylvania*

$9.95, cloth; $4.95, paper

EXPLORING CONTRADICTIONS
Political Economy in the Corporate State
Edited by Philip Brenner, *Trinity College*, **Robert Borosage and Bethany Weidner both of the** *Institute for Policy Studies*

"The collapse of the Gross National Product and the bankruptcy of all economic theory, never so apparent as today, is anticipated and penetrated in this collection of essays by some of the best brooding radical intelligences in the U.S."
—Hal Nieburg, *SUNY at Binghamton*

$7.95, cloth; $3.95, paper

DAVID McKAY COMPANY, INC.
750 Third Avenue, New York 10017

New from

New from the National Bureau of Economic Research

FOREIGN TRADE REGIMES AND ECONOMIC DEVELOPMENT

JAGDISH N. BHAGWATI and ANNE O. KRUEGER, Editors

This ambitious research project sponsored by NBER analyzes the phenomena of exchange control, liberalization, and economic development in less developed countries. There will be ten volumes published on individual countries plus a synthesis volume prepared by the co-directors, Jagdish N. Bhagwati and Anne O. Krueger.

The three major topics covered in each country studied are: the Anatomy of Control, the Liberalization Episode, and Growth Relationships.

The first two volumes in the series are:

Volume 1: Turkey, by Anne O. Krueger
cloth, $17.50; paper, $5.00
Volume 2: Ghana, by J. Clark Leith $15.00

A Special Conference of the National Bureau of Economic Research

ESSAYS IN THE ECONOMICS OF CRIME AND PUNISHMENT

GARY BECKER and WILLIAM LANDES

The distinguishing and unifying feature of the essays in this volume is the systematic study of enforcement as an economic problem. Subjects covered include optimal enforcement, optimal bail system, and the development of a positive theory of legal decision making as applied to enforcement decisions.

Human Behavior and Social Institutions, 3
cloth, $12.50; paper, $3.50

INCOME INEQUALITY

Regional Analyses Within a Human Capital Framework

BARRY R. CHISWICK

This human capital analysis focuses on differences in the amount of investment in schooling and in the rate of return from this investment to explain variations in inequality of labor income in the United States and other countries.

Human Behavior and Social Institutions, 4 $15.00

COLUMBIA UNIVERSITY PRESS

Address for orders: 136 South Broadway, Irvington, New York 10533

Columbia

PEOPLE OR PENGUINS

The Case for Optimal Pollution

WILLIAM F. BAXTER

Transcending the fuzzy thinking and emotional rhetoric associated with environmental discussion, Professor Baxter has provided an engagingly written assessment of the legal and economic questions which lie at the heart of all pollution problems. cloth, $6.00; paper, $1.95

THE FAILURE OF U.S. ENERGY POLICY

RICHARD B. MANCKE

Present energy policy was designed by officials who have misperceived our world, Dr. Mancke maintains in this scathing indictment. His analysis is bolstered by historical case studies and supplemented by reform programs for the future. cloth, $10.00; paper, $2.95

INTRODUCTION TO MANAGERIAL ECONOMICS

WILLIAM F. SHARPE

How much economic theory should the modern business manager know? Developed from ideas first presented in Sharpe's highly-praised *The Economics of Computers,* this book provides in concise form all the basics the student or working manager will require, and supplies the necessary background for more complex analyses.
cloth, $10.00; paper, $3.50

SEX, DISCRIMINATION, AND THE DIVISION OF LABOR

CYNTHIA B. LLOYD, Editor

This collection of sixteen essays looks at the role of women in the labor market, touching on the areas of productivity, participation, and compensation, among others. These essays will be valuable to the student or individual with only a basic knowledge of economics as well as to students of economic research and teachers of courses on women.

Studies in Economics, 8 cloth, $15.00; paper, $6.00

Ready in May for your fall courses.

Write today for examination copies for fall course adoption.

COLUMBIA UNIVERSITY PRESS

Address for orders: 136 South Broadway, Irvington, New York 10533

IASP TITLES IN ECONOMICS

John Kenneth Galbraith and the Lower Economics

Second, Expanded Edition

Myron E. Sharpe, Editor of CHALLENGE: The Magazine of Economic Affairs

"The best review I have yet seen of Galbraith's works . . . Sharpe (in the first edition) has succeeded in highlighting Galbraith's contribution to economic thought, while exposing to critical scrutiny the weakest parts of his analysis. I'm happy to recommend this book." **—Robert L. Heilbroner**

This edition contains a new section evaluating Galbraith's *Economics and the Public Purpose*, and includes a lengthy interview with Galbraith. $7.95

The Socialist Economies of the Soviet Union and Europe

Marie Lavigne, University of Paris

"Marie Lavigne's book is both thoroughly competent and of a sufficiently high standard to raise French Sovietology to the level of English or American works on the same subject." **—Revue Economique** **(Reprinted in Economic Journal)**

This encyclopedic study devotes major attention to the Soviet economic system, and is unusually detailed in describing the other socialist economies and in drawing upon them for important cross-national comparisons.

Dr. Lavigne's publications include *Capital in the Soviet Economy, The Problem of Prices in the Soviet Union* (with Prof. H. Denis), and *COMECON: From Cooperation to Integration of the Socialist Economies.* $20.00

Land Reform and Economic Development in China

A Study of Institutional Change and Development Finance

Victor D. Lippit, University of California at Riverside

A new analysis that makes a significant contribution to the economics of the Chinese Revolution and to development theory generally.

Professor Lippit shows how a poor country managed simultaneously to raise the share of national income devoted to investment and increase living standards. He focuses on the use of the "surplus" — the capital or property share of national income — to redirect income flows, thereby to shift consumption from luxuries to necessities and increase public investment. $15.00

International Arts & Sciences Press Inc.
901 North Broadway
White Plains, New York 10603

Essays in Honor of Jan Tinbergen

Edited by Willy Sellekaerts, The Wharton School, University of Pennsylvania

Three volumes of original articles by some of the world's leading economists. Each volume includes a selected bibliography of Tinbergen's work, as well as an introduction which appraises Tinbergen's major contributions to economics. The three volumes are: *International Trade and Finance; Economic Development and Planning; Econometrics and Economic Theory.* $20.00 per volume

The Strategy of International Development

Essays in the Economics of Backwardness

H. W. Singer

Edited by Sir Alec Cairncross and Mohinder Puri

This volume contains H. W. Singer's most important papers of the past five years on the central problems of economic development. The collection presents the author's current views after more than twenty-five years of study and direct involvement with many of the developing countries.

Sir Alec Cairncross has written a substantive introduction that reviews Dr. Singer's distinguished career and examines in some detail the role of technological development in economic growth. $20.00

Self-Governing Socialism: A Reader

Edited by Branko Horvat, Mihailo Marković and Rudi Supek

Three distinguished Yugoslav scholars have selected seminal articles, book extracts, and documents from 20 countries in order to trace the theory and practice of self-governing socialism from the early nineteenth century to the present. Much of the material is being made available in English for the first time. Volumne I: *Historical Development; Social and Political Theory.* Volume II: *Sociology and Politics; Economics.*

Vol. I (cloth)	$21.50	Vol. II (cloth)	$18.50
(paper)	$9.50	(paper)	$7.95

Monetary Cooperation Between East and West

Adam Zwass, Vienna Institute for Comparative Economic Studies

Foreword by George Garvy

A systematic study of the instruments and processes of East-West financial relations based on a critical analysis of the money, credit and foreign trade systems of the East European countries and of their Council for Mutual Economic Assistance.

Dr. Zwass occupied managerial positions in the central banking systems of Poland and the USSR. From 1963 to 1968 he was a Counsellor in the COMECON Secretariat in Moscow, where he was responsible for the work of the International Bank for Economic Cooperation. $15.00

ELSEVIER

SCIENTIFIC PUBLISHING COMPANY, INC. 52 VANDERBILT AVENUE, NEW YORK, NEW YORK 10017

Psychological Economics

By GEORGE KATONA.

Consumer behavior is shown to be of crucial importance in today's economy. During the last few decades in all highly developed economies, consumers have acquired great discretion in spending and saving. Changes in their attitudes and expectations, determined through sample interview surveys, help us to understand and predict upswing or downswing in the economy as well as the extent of inflation. The book's timeliness is exemplified in the coverage of such problems as the psychology of inflation and the possible limits to growth. **1975. 352 pages. $13.00**

Games for Society, Business and War

TOWARDS A THEORY OF GAMING

By MARTIN SHUBIK.

Designed primarily for those interested in gaming, this original and timely book is an overall survey in which the difference between gaming and the theory of games is clearly distinguished. It provides a unified view of both the theory and usage of gaming for teaching, research, business, military science, social psychology and other purposes. Chapters discuss and illustrate such topics and areas as: the revelance of simulation, behavioral models of systems, mathematical and game theory models of conflict and cooperation. **1975. 380 pages. $16.50**

The Uses and Methods of Gaming

By MARTIN SHUBIK.

A companion volume to *Games for Society, Business and War*, this guide presents information on the many types and purposes of gaming. Designed for those who wish to construct games, it stresses how to set up, to cost, and to run games. In addition, it tells where to find the basic literature and information on gaming and related topics. **1975. 208 pages. $12.00**

Human Behavior in Economic Affairs

ESSAYS IN HONOR OF GEORGE KATONA

Editors: BURKHARD STRUMPEL, JAMES N. MORGAN, and ERNEST ZAHN.

George Katona is renowned as the pioneer in the development of psychological economics and the 25 essays assembled here are written by prominent people in related fields. The essays evaluate and review the progress and success of the behavioral economic science, stressing the power of an interdisciplinary approach to the understanding of the behavior of the householder as consumer. **1972. 590 pages. $18.75**

Efficiency in Universities

THE LA PAZ PAPERS

By KEITH LUMSDEN.

The crucial question of how university efficiency can be improved is dealt with in this volume. In practice, the management of universities is carried out by administrators who depend largely on government funding and/or private donations, and each tries to use the university as a means of furthering its own objective. Various other facets are treated in this volume, including defining different functions of the university, improving efficiency through reorganization and the introduction of new techniques. **1974. 278 pages. $15.95**

Advertising Planning

MATHEMATICAL MODELS IN ADVERTISING MEDIA PLANNING

By DENNIS H. GENSCH. **1973. 176 pages. $12.50**

For more information about these books, write to Department Z at the address above.

ORGANIZATION FOR ECONOMIC COOPERATION AND DEVELOPMENT

Headquarters: 2, rue André-Pascal, 75775 Paris Cedex 16, France

OECD

economic growth
development aid
trade expansion

Periodical Aids for Reference, Teaching, and Planning	Subscription Rates *Surface*	Subscription Rates *Air*
ECONOMIC OUTLOOK. A periodic assessment of economic trends and prospects in the main trading nations, published each July and December, $5 each. *"Occasional Studies,"* which treat selected topical problems in some depth, are included, or may be obtained separately ($2.50 each). Combined annual subscription:	$ 8.50	$ 10.50
OECD FINANCIAL STATISTICS. A collection of data on financial markets in 15 European countries, the United States, Canada and Japan. Of particular interest to investment trusts and other financial institutions. Issued yearly, with periodic supplements.	37.50	65.50
MAIN ECONOMIC INDICATORS. A unique tabulation of international indicators of economic activity. Graphics emphasize important changes in key indicators, and quarterly supplements provide additional data on consumer prices and industrial production. Monthly. Single copy, $2.50.	27.50	51.50
ECONOMIC SURVEYS. Annual review of each OECD country. Individual surveys, $1.75 each. Entire series:	22.50	41.50
PROVISIONAL OIL STATISTICS. Crude oil, gas, fuels. Output, trade and bunkers; by quarters. 4 issues yearly.	4.50	6.50
GUIDE TO LEGISLATION ON RESTRICTIVE BUSINESS PRACTICES IN EUROPE, NORTH AMERICA AND JAPAN. Includes a selection of relevant court and administrative decisions, up-dated by periodic supplements. Supplements, $10 per series, $16.00 by air. Basic volume:	50.00	
FOREIGN TRADE STATISTICS		
SERIES A. Overall trade by countries. Now monthly, plus half-yearly supplements.	30.00	45.00
SERIES B. Commodity trade, analysed by main trading areas. 6 volumes quarterly.	30.00	79.00
SERIES C. Detailed analysis by products. Imports and exports. Semi-annually.	35.00	107.00
LIAISON BULLETIN. Summarizes research under way in research and training institutes throughout the world. 4 issues yearly. Directories for Asia, Africa, Latin America and Europe available separately ($2.00 each).	8.50	14.00
NUCLEAR LAW BULLETIN. Legislative and regulatory activities, case law and administrative decisions, international organizations and agreements. 2 issues yearly, plus supplements.	4.50	7.30
PULP AND PAPER. Quarterly data on stocks of paper woodpulp, the production and trade of pulp and paper and board, the shipments of market pulp. Each issue, $3. Annually:	$9.00	
RESEARCH ON TRANSPORT ECONOMICS. Six-monthly information bulletin containing a subject index and an index of organizations mentioned. 2 issues yearly, May and November. (Sold only by annual subscription)	$28.00	
OECD OBSERVER. Illustrated bi-monthly magazine selectively covering the full range of OECD activities, and including announcements of new OECD publications. Sample copy on request. 6 issues yearly.	4.50	7.50

Rates subject to change.

New Report on Energy

ENERGY PROSPECTS TO 1985.

An Assessment of Long Term Energy Developments and Related Policies. **$11.25**

MEMBER COUNTRIES

Australia
Austria
Belgium
Canada
Denmark
Finland
France
Germany
Greece
Iceland
Ireland
Italy
Japan
Luxembourg
Netherlands
New Zealand
Norway
Portugal
Spain
Sweden
Switzerland
Turkey
United Kingdom
United States

SPECIAL STATUS Yugoslavia

OECD also publishes about 100 monographs yearly.

A Blanket Order Plan is available.

Send orders and requests for information to:

OECD PUBLICATIONS CENTER
SUITE 1207E 1750 PENNSYLVANIA AVE., N.W.
WASHINGTON, D.C. 20006 • (202) 298-8755

The American Economic Review

ARTICLES

E. S. MILLS AND F. M. PETERSON
Environmental Quality: The First Five Years

E. F. FAMA — Short-Term Interest Rates as Predictors of Inflation

J. E. STIGLITZ — The Theory of "Screening," Education, and the Distribution of Income

P. J. LLOYD — Substitution Effects and Biases in Nontrue Price Indices

M. K. BLOCK AND J. M. HEINEKE
A Labor Theoretic Analysis of the Criminal Choice

R. A. MEYER — Monopoly Pricing and Capacity Choice Under Uncertainty

M. N. BAILY — Dynamic Monopsony and Structural Change

D. A. WISE — Academic Achievement and Job Performance

L. R. CHRISTENSEN, D. W. JORGENSON, AND L. J. LAU
Transcendental Logarithmic Utility Functions

R. B. WESTIN — Empirical Implications of Infrequent Purchase Behavior in a Stock Adjustment Model

I. EHRLICH — The Deterrent Effect of Capital Punishment: A Question of Life and Death

L. G. REYNOLDS — China as a Less Developed Economy

P. S. HELLER — A Model of Public Fiscal Behavior in Developing Countries: Aid, Investment, and Taxation

SHORTER PAPERS: E-M. Claassen; M. Allais; J. L. Scadding; I. Friend and R. Westerfield; G. J. Benston; M. E. Canes; W. Hamovitch and R. D. Morgenstern; D. A. Katz; J. Barron; C. C. Holt; G. Hanoch; W. L. Holahan; W. A. Bomberger; S. A. Ross and M. L. Wachter; G. Smith; D. R. Bohi and G. W. Scully; G. J. Stigler.

JUNE 1975

THE AMERICAN ECONOMIC ASSOCIATION

Founded in 1885

- Published at George Banta Co., Inc., Menasha, Wisconsin.

- The American Economic Review, including four quarterly numbers, the *Proceedings* of the annual meetings, the *Directory*, and Supplements, is published by the American Economic Association and is sent to all members five times a year, in March, May, June, September, and December.

- Membership dues of the Association are $23.00 a year, which includes a year's subscription to both the *American Economic Review* and the *Journal of Economic Literature*. Junior membership at $11.50 per year is available to registered students for a maximum of three years. Subscriptions (libraries, institutions, or firms) are $34.50 a year, and only subscriptions to both publications will be accepted. Please write the Secretary's office, Nashville, Tennessee for single copy prices.

- Correspondence relating to the *Papers and Proceedings*, the *Directory*, advertising, permission to quote, business matters, subscriptions, membership and changes of address should be sent to the Secretary, Rendigs Fels, 1313 21st Avenue South, Nashville, Tennessee 37212. Change of address must reach the Secretary at least six (6) weeks prior to the month of publication. The Association's publications are mailed second class and are not forwarded by the Post Office.

- Second-class postage paid at Nashville, Tennessee and at additional mailing offices. Printed in U.S.A.

Officers

President

ROBERT AARON GORDON
University of California, Berkeley

President-Elect

FRANCO MODIGLIANI
Massachusetts Institute of Technology

Vice-Presidents

MARTIN BRONFENBRENNER
Duke University
CHARLES L. SCHULTZE
University of Maryland

Secretary-Treasurer and Editor of the Proceedings

RENDIGS FELS
Vanderbilt University

Assistant Secretary-Treasurer

C. ELTON HINSHAW
Vanderbilt University

Managing Editor of The American Economic Review

GEORGE H. BORTS
Brown University

Managing Editor of The Journal of Economic Literature

MARK PERLMAN
University of Pittsburgh

Executive Committee

Elected Members of the Executive Committee

IRMA ADELMAN
University of Maryland
ROBERT HEILBRONER
New School for Social Research
ANDREW F. BRIMMER
Harvard University
PAUL M. SWEEZY
Monthly Review
CAROLYN SHAW BELL
Wellesley College
BURTON A. WEISBROD
University of Wisconsin, Madison

Ex Officio Members

KENNETH J. ARROW
Harvard University
WALTER W. HELLER
University of Minnesota

THE AMERICAN ECONOMIC REVIEW

GEORGE H. BORTS
Managing Editor

WILMA ST. JOHN
Assistant Editor

Board of Editors

BELA BALASSA
EUGENE F. FAMA
MARTIN S. FELDSTEIN
ANN F. FRIEDLAENDER
STEPHEN M. GOLDFELD
ROBERT J. GORDON
ROBERT E. HALL
BENT HANSEN
JAMES R. MELVIN
WILLIAM D. NORDHAUS
LEONARD A. RAPPING
STEPHEN RESNICK
ANNA J. SCHWARTZ
JEROME STEIN
JOSEPH E. STIGLITZ
S. C. TSIANG
FINIS WELCH
MARINA V. N. WHITMAN

- Manuscripts and editorial correspondence relating to the regular quarterly issue of this REVIEW should be addressed to George H. Borts, Managing Editor of THE AMERICAN ECONOMIC REVIEW, Brown University, Providence, R.I. 02912. Manuscripts should be submitted in duplicate and in acceptable form and should be no longer than 50 pages of double-spaced typescript. A $15 submission fee must accompany each manuscript. *Style Instructions* for guidance in preparing manuscripts will be provided upon request to the editor.

- No responsibility for the views expressed by authors in this REVIEW is assumed by the editors or the publishers, The American Economic Association.

- Copyright American Economic Association 1975.

June 1975

VOLUME LXV, NUMBER 3

CALCUTTA

Articles

Environmental Quality: The First Five Years
Edwin S. Mills and Frederick M. Peterson 259

Short-Term Interest Rates as Predictors of Inflation *Eugene F. Fama* 269

The Theory of "Screening," Education, and the Distribution of Income *Joseph E. Stiglitz* 283

Substitution Effects and Biases in Nontrue Price Indices *P. J. Lloyd* 301

A Labor Theoretic Analysis of the Criminal Choice
M. K. Block and J. M. Heineke 314

Monopoly Pricing and Capacity Choice Under Uncertainty *Robert A. Meyer* 326

Dynamic Monopsony and Structural Change
Martin Neil Baily 338

Academic Achievement and Job Performance
David A. Wise 350

Transcendental Logarithmic Utility Functions
Laurits R. Christensen, Dale W. Jorgenson, and Lawrence J. Lau 367

Empirical Implications of Infrequent Purchase Behavior in a Stock Adjustment Model
Richard B. Westin 384

The Deterrent Effect of Capital Punishment: A Question of Life and Death *Isaac Ehrlich* 397

China as a Less Developed Economy
Lloyd G. Reynolds 418

A Model of Public Fiscal Behavior in Developing Countries: Aid, Investment, and Taxation
Peter S. Heller 429

Shorter Papers

Demand for International Reserves and the Optimum Mix and Speed of Adjustment *Emil-Maria Claassen* 446
The Hereditary and Relativistic Formulation of the Demand for Money: Circular Reasoning or a Real Structural Relation? *M. Allais* 454
Professor Allais' Theory of the Demand for Money: Rejoinder *John L. Scadding* 465
Required Disclosure and the Stock Market:
  Comment *Irwin Friend and Randolph Westerfield* 467
  Rejoinder *George J. Benston* 473
The Simple Economics of Incentive Contracting: Note *Michael E. Canes* 478
The Principal Cause of Academic Salary Differentials: Research Output or Experience?:
  Comment *William Hamovitch and Richard D. Morgenstern* 484
  Reply *David A. Katz* 486
Segmentation of the Labor Market:
  Comment *John M. Barron* 487
  Rejoinder *Charles C. Holt* 491
The Elasticity of Scale and the Shape of Average Costs *Giora Hanoch* 492
The Welfare Effects of Spatial Price Discrimination *William L. Holahan* 498
Wage Determination, Inflation, and the Industrial Structure:
  Comment *William A. Bomberger* 504
  Reply *Stephen A. Ross and Michael L. Wachter* 507
Pitfalls in Financial Model Building: A Clarification *Gary Smith* 510
Buyer's Prices, Seller's Prices, and Price Flexibility:
  Comment *Douglas R. Bohi and Gerald W. Scully* 517
  Reply *George J. Stigler* 526

Errata 527

Notes 529

Environmental Quality: The First Five Years

By EDWIN S. MILLS AND FREDERICK M. PETERSON*

On January 1, 1970, President Nixon signed the National Environmental Policy Act (*NEPA*). Explicitly modeled on the Employment Act of 1946, *NEPA* established the Council on Environmental Quality (*CEQ*) in the president's executive offices to advise the president on environmental issues as the Council of Economic Advisers (*CEA*) advises him on economic issues. Just as *CEA* is required to submit to the president, *The Annual Report of the Council of Economic Advisers* (*ER*), so *CEQ* must submit an annual report, *Environmental Quality* (*EQ*).

Each year since 1971, this *Review* has devoted part of an issue to comments on *ER*. But despite the substantial and growing interest of economists in environmental problems, the profession has taken little notice of successive volumes of *EQ*. The purpose of this paper is to use the occasion of the publication of the fifth volume of *EQ* to bring this important publication to the attention of the economics profession. We start with a brief comparison between *CEQ* and *CEA*, then present an overview of *EQ*, and finally a review of its economic analysis.

I. Organizational Background

CEQ and *CEA* are in many ways similar organizations. Each consists of three members appointed by the president, with one member designated chairman. Each is served by a staff that includes about twenty-five professionals. Each not only advises the president, but also advises, cajoles, and intermediates among operating agencies of the federal government. There are two important differences between *CEQ* and *CEA*.

First, whereas *CEA* and its professional staff are, quite properly, almost all economists, *CEQ* and its staff are, also properly, drawn from a variety of professional backgrounds. Nevertheless, an economist is struck by the paucity of economic input at *CEQ*. No economist was among those appointed by President Nixon to *CEQ* and they have always been a tiny minority on its staff. Although both *CEA* and *CEQ* have had rapid and probably desirable turnover of their members and staffs, many *CEQ* personnel have pursued large parts of their careers in government whereas many *CEA* personnel are on leave from universities and research institutions.

Second, whereas *CEA* is entirely a staff organization, *CEQ* has an important operational responsibility. The act (*NEPA*) instructed all federal agencies to " . . . include in every recommendation or report on proposals for legislation and other major Federal actions significantly affecting the quality of the human environment, a detailed statement by the responsible official on—(i) the environmental impact of the proposed action. . . . " Congress must have written this directive while distracted by other matters. It has led to endless litigation—so much so that it is referred to as the lawyers' relief bill—stemming from the failure of Congress to specify how the impact statements should relate to other laws that incorporate environmental considerations into decisions;

---

* Princeton University and the University of Maryland, respectively.

who if anybody should receive and read the statements; and what, if any, effect the statements should have on decisions taken. In any case, it has fallen to *CEQ* to evaluate impact statements. This task now occupies the time of a substantial minority of *CEQ* staff. Indeed, tracing and disentangling *NEPA* and its impact statements have occupied a substantial part of the five volumes of *EQ*.

II. An Overview of *Environmental Quality*

EQ and *ER* are intended to serve three purposes. First, they provide a systematic survey of trends and conditions regarding the environment and the economy. The first chapter of *ER* always surveys trends in the economy's output, employment, prices, etc. Likewise, each issue of *EQ* contains a chapter that discusses recent trends in air and water pollution, solid waste disposal, etc.

Second, they serve the educational purpose of providing the informed public with a systematic and coherent analysis of problems of urgent social concern. *ER* frequently presents analyses of causes and consequences of economic problems and of the pros and cons of alternative government policies to deal with the problems. Aside from the perennial problem of economic stabilization, frequent candidates for analysis are balance of payments, agriculture, poverty, and industrial organization and regulation. In the same vein, *EQ* provides frequent analyses of environmental and related problems. There are annual discussions of government policy developments toward air and water pollution and solid waste disposal. There are frequent sections on pesticides, radiation, and noise, and occasional sections on energy and natural resource adequacy. In addition, *EQ* sometimes presents a case study of a particularly instructive attempt to clean up a river or estuary.

Third, *EQ* and *ER* are intended to present and defend the president's program in their areas of concern. In *ER*, this purpose is served by methods varying from hints and suggestions on future directions for fiscal and monetary policy to full chapters laying out strategy for the war on poverty. In *EQ*, the third purpose has been of decreasing importance. The first volume (1970) had a concluding chapter that laid out boldly a list of environmental policy "needs." The second volume (1971) had a competent and favorable discussion of the advantages of effluent fees as a tool of environmental policy. But the last three volumes have had almost no discussion of alternative policies, except occasional favorable comments on specific bills previously proposed by the president. One gets the strong impression that *EQ* has become very much the status report of *CEQ* and very little the policy statement of the administration.

The first purpose is carried out in an almost entirely nonpolitical way in both *ER* and *EQ*. Despite occasional slips, *ER* and *EQ* have reported past trends with remarkable objectivity and with remarkably little attempt to put one administration or another in a particularly good or bad light.

The second purpose tends to be executed with considerable delicacy, but also with considerable objectivity. Usually, the issue seems to be how specific the professionals who write the reports are permitted to be by the administration regarding conclusions for government policies to solve problems. This sometimes causes analyses and conclusions to be vague or evasive and studded with circumlocution, but examples of blatantly partisan analysis of policy problems are remarkably rare in both *ER* and *EQ*. Partisan-minded readers can, however, see considerable political influence in choice of topic. The best example of this is the treatment of poverty in *ER*. A long section or chapter on poverty had become *de rigueur* in the Democratic administrations' *ER*'s starting in 1964. But the first discussion of poverty in the five *ER*'s submitted to President Nixon appeared in his last *ER* (1974). Partisan influences on choice of topic are harder to identify in *EQ* since we have not yet had a Democratic report. But the five Republican *EQ*'s have devoted considerable attention to initiatives of state and local governments to clean up the environment. It seems certain that this attention has been inspired more by the administration's desire to deemphasize the federal role than by the meagre initiatives reported, and that state and local initiatives would have received less attention in *EQ*'s produced by Democratic administrations.

The third purpose is of course explicitly political, and partisanship is to be expected. In fact, *EQ* has been remarkably objective in discussing the president's environmental programs and his frequent battles with Congress on the subject. The best example of this contention is the controversy surrounding the 1972 amendments to the Water Pollution Control Act. This was certainly the most important environmental legislation of the Nixon presidency and was passed over his veto. Although he agreed with many of its complex substantive provisions, he felt the large sums it provided for municipal waste treatment plant construction were inflationary. After it was passed over his veto, he proceeded to impound large parts of its appropriations and there followed an intense political and legal battle. The 1973 *EQ* has an extended factual discussion of these matters and contains no hint of recrimination or partisanship.

Economists accustomed to *ER* inevitably feel that *EQ* is long and diffuse. In fact, *EQ* is a long report. Its 1974 issue was 597 pages, two-thirds longer than the 1974 *ER*. Both *EQ* and *ER* grew during the Nixon years, but *EQ* grew more rapidly. The fifth issue of *EQ* was 82 percent longer than the first. In part, *EQ* is long because of the broad definition *CEQ* has given to the term environment. When economists discuss environmental problems they invariably refer to the effects of materials discharges in the course of extraction, fabrication, and consumption, on the quality of the ambient air and water environments. Laymen and other professionals sometimes use the term more broadly, and *CEQ* appears to have adopted the broadest possible definition. They have discussed the adequacy of fuel and other natural resource reserves as well as effects of their extraction and combustion on air and water environments; the urban environment, including aesthetic effects of building size, design and location, the functions of downtowns, and urban renewal; urban land use, including effects of urban decentralization or sprawl; national parks; population problems; and details regarding the organization and administration of the federal government's environmental program. The effect on *EQ* is to make the report long and confusing.

The effect of the broad definition of the subject would not be so serious if *EQ* were better organized. The minimum that the reader ought to be able to expect from *EQ* is a clear and concise presentation of environmental trends and status. There should be one chapter or section each for air, water, and solid wastes, and they should appear consistently near the front of the report. Each should discuss the status and trends of discharges, ambient environmental quality, and effects of pollution on people and property. Great pains should be taken each year to update and improve an intelligible and meaningful set of quantitative measures of discharges and of ambient quality.

In fact, the first (1970) *EQ* started off to do just that. But in the second (1971) volume, one had to go to chapter 7 for such data. In the third (1972) volume, the presentation of comprehensive data was abandoned altogether, and a discussion of uninformative environmental indices was substituted. The fourth (1973) volume reverted, in chapter 6, to a report on trends. The indices discussed the previous year were blessedly abandoned, but much of the ambient data were presented as ratios to federal standards. The fifth (1974) volume presented, in chapter 3 this time, a good discussion of airborne discharges and ambient air quality, but almost entirely in graphs rather than tables, making it difficult to read precise values. The water data are much poorer despite the fact more water than air data are available.

The presentation and discussion of data should be followed by a discussion of policy changes and developments with regard to air, water, and solid waste. If this format were followed, each report could then also include one or more chapters on some of the special topics mentioned above, especially if *CEQ* had something to say about their relationship to air and water quality. Instead, the central chapters on environmental trends have been located in every part of *EQ* and, in some reports, the special topics have been made the focus of the report.

Closely related to the above discussion is the issue of data. One of the charges to *CEQ* in *NEPA* is to improve the collection and publication of data on environmental quality. Everybody knows that data on environmental quality are inadequate, and we do not want to add to the chorus of complaints on the subject. But the fact is that data on environmental quality have improved greatly during the last five or so years, and *U.S.* data are as good and plentiful as those in almost any country. Rather, our complaint is that *EQ* should have started at the beginning to present, preferably in a statistical appendix, summaries of available data, updating the series as new data became available. They should have devoted considerable effort to presenting data in consistent and informative ways. They should explain what the data mean, how they are compiled, where additional detail can be found, and how the data can be compared when series are changed or expanded. Finally, *EQ* should indicate gaps and inadequacies in the data.

ER has of course contained a statistical appendix for years. Although the original sources of the data are not dispersed or obscure, the *ER* appendix is invaluable as a ready reference for overall economic trends. Environmental data, in contrast, are dispersed and inaccessible. A good statistical appendix to *EQ* would be even more valuable than that in *ER*. The most important data to present are averages of discharges and ambient air and water quality. Data should be neither excessively detailed nor excessively aggregated. The most detailed data can be obtained from the original sources and do not need to be published in *EQ*. If *CEQ* undertakes or supports studies of environmental indices, they can be presented, but they should be in addition to the more basic data rather than instead of them.

The 1974 *EQ* makes a belated and inadequate attempt at an environmental statistical abstract. Of the thirty tables presented, only the last five pertain to discharge quantities or to ambient air and water quality. The others, like the texts of *EQ*, are a miscellany of data on agriculture, minerals, energy, population, wholesale prices, etc. There is little harm in including some of this information, but it is presented without organization and with no apparent thought as to what to include. Much more important, of the five tables relating to pollution, not one provides direct measures of ambient air or water quality. We are given only numbers of stations at which ambient air and water quality did or did not meet standards set by the government. It is shocking that the ambient air and water quality data presented in *EQ*'s statistical appendix are even worse than those presented in the *Statistical Abstract*.

Lest it be thought that we are setting a standard for *EQ* that is theoretically desirable but practically unattainable, we refer briefly to the analogous Japanese publication, *Quality of the Environment in Japan*. Japan's basic environmental data are certainly no better than those available in the *U.S.* The Japanese passed their first comprehensive antipollution laws long after we passed ours and they have had less experience in writing their annual environmental report than we have. Yet their report is only half as long as ours (at least in its English language version), less cluttered with peripheral issues, less redundant, better organized and better documented with intelligible and meaningful data. We do not set the Japanese report as an ideal. There are many things wrong with it, but we do not think it unreasonable to expect the *U.S.* report to be as good as the Japanese. Yet it is not.

III. Economic Analysis in *Environmental Quality*

We come now to the central issue, the kinds of economic analysis in *EQ*. Economics is of course only one part of the process of government policy making toward the environment. The metering of discharges and of ambient environmental quality, the measurement of damages done by pollutants, the analysis of physical measures to abate discharges and to reduce their polluting effect on the environment, and the analysis of the environmental effects of a wide range of government activities all require the work of physical and biological scientists and of engineers.

Yet environmental problems and remedial government policies cannot be analyzed satisfactorily without a substantial input of economic analysis. We find it useful to classify economic analyses of environmental problems into three categories. First is the study of market failure, why and in what circumstances private production and consumption of goods and services result in excessive or excessively harmful discharges of materials to the environment. Second is benefit-cost studies of varying amounts of pollution abatement. Third is government program evaluation. All government programs to improve environmental quality entail complex interactions between government and private sectors. Each possible program has peculiar effects on the ways private markets allocate resources, and such effects are difficult to foresee without careful analysis. In addition, some government programs are much easier or cheaper to administer than others, and provide varying incentive structures for the agencies that administer them. Analysis of these issues is somewhere around the fuzzy borderline between economics and public administration.

The tools in the economist's box are of course by no means perfect for these purposes and it is now fashionable for economists to beat their breasts about their inadequacy. Much of this is overdone. The basic reasons for market failure regarding polluting discharges were well known to Pigou and have become even better understood as a result of sharper analysis in the last decade or two. Benefit-cost analysis, although currently out of favor in the national government, has been developed into a sophisticated tool of evaluation. Program evaluation remains more an art than a science, but a great deal of research has been undertaken about the efficacy and implications of alternative government programs to induce private decision makers to abate polluting discharges. Our overall evaluation is that environmental problems have been better studied and are better understood than most micro-economic public policy issues. We first evaluate the use made of economic analysis in *EQ* to shed light on government antipollution policy. We then conclude the section with brief comments on the use of economic analysis to understand other problems discussed in *EQ*.

The space devoted to economic analysis in the first five volumes of *EQ* is extremely small. For a publication that has printed the text of *NEPA* five times, printed and reprinted executive orders and administrative guidelines, traced the development of relatively minor changes in government programs in endless detail, and sometimes discussed the same topics and data in two or three places in one volume, the economic analysis is remarkably brief and isolated. (The only exception to this is the repeated discussion of pollution control costs, which we refer to below.) It is almost as though economists had been given a quota of a few pages in each volume, and that other parts of the volume, especially those surveying the development of government policies, had been written by others without reference to what the economists wrote. Despite this, *EQ* contains some good micro-economic analysis, certainly on a par with most of that contained in *ER*. In fact, there is a distinct and rather logical development of the economic analysis in successive volumes of *EQ*.

The first chapter of the 1970 volume of *EQ* contains the best overall picture of environmental problems we have seen in a government report published up to that time. This chapter contains a competent discussion of the basic economic cause of pollution, market failure because polluting discharges are an external diseconomy. But the discussion is less than one page! By contrast, the chapter is followed by a 10-page description of the organization of the federal government for pollution control, and then by more than 100 pages on the physical and biological aspects of pollution. No indication is given in these pages how economics might be used to shed light on the pollution problems discussed, on the costs and benefits of solving them, or on the efficacy of possible government programs to do so.

The 1971 volume of *EQ* contains a much more elaborate discussion of basic environmental economics, this time in chapter 4. It starts with a careful discussion of market failure regarding the environment, emphasizing the failure of usual market incentives properly to value common property resources—the ambient air and water environments. It then presents a frank and thorough survey of what little is known about the benefits of pollution abatement. This is followed by a much longer survey of the richer body of evidence about the costs of pollution control. Both the benefit and cost surveys cover government and private estimates objectively, carefully distinguishing among average and marginal costs and benefits. They conclude with a comparison between costs and benefits, suggesting where they think the balance lies, but frankly admitting the inability of existing studies to sustain a conclusion about the optimum amount of abatement. Finally, they present a remarkably good evaluation of alternative pollution abatement strategies. They rightly concentrate on discharge standards and effluent fees as the two most important candidates for pollution control strategies. They do not come down nearly as hard on the side of effluent fees as most academic economists do, but their conclusions are as firm as can be expected given that national policy depends entirely on discharge standards. They advocate a balanced use of standards and fees. This is to be applauded as a piece of bravery within the national administration, but as a policy, it would probably achieve the worst of both worlds.

The 1972 volume of *EQ* contains a chapter (8), devoted entirely to the costs of pollution control. It expands and updates the cost data in the 1971 volume. *CEQ*'s survey takes only about twelve pages. At that point the chapter is broken and the remainder of the chapter is a reprint of the overview of a large interagency study of the economic impact of pollution control. This produces an unsatisfactory discontinuity in the text. The studies summarized consist of micro studies of eleven important industries and a macro study. The micro studies analyze the effects of pollution control requirements on prices, investment, output, employment, and international competitiveness during the period 1972–76 in the eleven industries. It is hard to believe the results are very accurate in detail, but the overview is probably correct in concluding that the pollution control requirements then in effect would not cause very serious dislocations in most industries. The macro study includes public and private pollution control expenditures in a macro-econometric model and concludes, predictably, that their effects can be offset by changes in fiscal and monetary policy. Practically all the cost analyses in the 1972 volume of *EQ* were rendered obsolete by the passage of the 1972 amendments to the Water Pollution Control Act a few weeks after the report appeared.

The 1973 volume again includes a full chapter (3) of economic analysis. This time, as in 1971, both benefits and costs of pollution control are included. The conceptualization of the benefit-cost analysis is remarkably sophisticated. On the benefit side, the distinction is made between reductions in damages from pollution and reductions in costs of avoiding pollution damages. The data are still poor in comparison with available cost data, but the survey is competent. On the cost side, the distinction is made between abatement costs and transaction costs (the costs of administering pollution control programs), and the cost estimates are updated to take account of the more stringent requirements of the 1972 water quality amendments. In this volume, in contrast with the 1971 volume, no attempt is made to put the cost and benefit estimates together to estimate the abatement level at which they might be equated at the margin. It seems clear that the elaborate 1973 cost estimates were undertaken defensively, in response to growing industry complaints that the 1970 air and 1972 water amendments would impose large dislocations on the economy. Hence, *CEQ*'s interest was not so much in comparing benefits and costs as in showing that the economy could afford the costs. Unfortunately, the 1973 volume contains no program evaluation. A short section on economic incentives merely points out a few of many undesirable incentive effects of our regulatory program. But there is no discussion of effluent fees or of other alternatives to the regulatory program.

Unfortunately, economic analysis of pollution problems is nearly absent from the 1974 *EQ*. We are given only a five-page update in chapter 2 of estimated abatement costs during the 1973–82 period. These most recent estimates are that federal antipollution programs cost the *U.S.* economy $6.3 billion in 1973 and that existing legislation will impose an average cost of about $20 billion per year on the economy during the 1973–82 period. *EQ* is content merely to point out that the former figure is about .75 percent of the 1973 *GNP* and the latter about 1 percent of the average projected *GNP* during the decade. But $20 billion per year is a lot of money, and it now becomes extremely important to ask whether these expenditures are justified by comparable benefits and whether the benefits could be obtained more cheaply by alternatives to existing programs. Thus it is extremely unfortunate that *CEQ* has backed off from benefit estimation and program evaluation when we are clearly moving onto the rapidly rising part of the marginal abatement cost curve.

We evaluate *EQ*'s environmental economic analysis under the following four headings.

1. In one volume or another, *EQ* has presented as good a menu of economic analyses of environmental problems as can be expected in the short space devoted to it. But the space allocated to economic analysis is too small to do justice to the complex issues involved. The severe space constraints can hardly be explained by overall space constraints since space is treated like a nearly free good in most of the report.

2. The most serious criticism of the economic analysis in *EQ* is that it is isolated from the rest of the report. For example, the 1971 volume presents a good section on program evaluation, but there is no evidence the analysis there presented has had any effect on the interminable histories of government policy changes that have been presented. Clearly, the section on program evaluation should establish the framework within which program development will be discussed. Instead, *EQ* treats economists as an interest group which must be mollified, as many interest groups are mollified in government reports, by including a section on a subject dear to their hearts.

3. We are disturbed at the trend in recent volumes to restrict economic analysis to whether the economy can "afford" the government's program of pollution abatement. This is not a very challenging economic issue. There can be no doubt that the *U.S.* economy can afford $20 billion, or $50 billion, per year on pollution abatement, provided urgent benefits to health and welfare are contingent on the expenditures. The important issues, for the public and for economists, are whether the costs are justified by comparable benefits and whether existing programs are the most efficacious ways to obtain the benefits.

4. Closely related to the previous point is the tendency to slight benefit estimation. *CEQ* and other federal agencies concerned with the environment have lavished large sums of money on pollution abatement cost studies. But the benefit side has been sadly neglected. Benefit studies are of course more difficult than cost studies, but that is justification for more rather than fewer resources devoted to them since public policy cannot possibly be set without explicit or implicit benefit estimates. Some careful benefit studies have been undertaken, especially regarding auto pollution control, in recent years, but they have come from universities, private research institutions, and the National Academy of Sciences. The tendency in government and, sad to say, sometimes among economists is to claim that economics can be applied to the cost side of pollution control but that the benefit side is necessarily a "political issue." In a democracy, both benefit and cost estimates are quite properly cast into law and administrative procedures by the political process. The purpose of expert advice is to edify the process. There is no reason to believe that economists can be of less help on the benefit than on the cost side.

It was mentioned in Section II that many subjects in addition to pollution are addressed in *EQ*. We conclude this section with brief comments on the use of economics in analyzing these issues. For some of these issues, it is quite justified that the analysis contains no economics. For example, each volume of *EQ* contains some discussion of international aspects of pollution. Subjects discussed include polluting discharges that move across national boundaries, pollution of the oceans, inadvertent modification of climate by man's activities, and cooperation among governments on environmental problems. Despite some recent work, there is at present relatively little that economists have to say on most of these subjects.

Land use, however, is a different matter. Each of the five volumes of *EQ* has a chapter or major section on some aspect of land use. Sometimes the subject is broadly construed, sometimes more narrowly in terms of urban or inner city land use. Sometimes the discussion is in terms of the intriguing relationships between pollution and land use, but more often it is in terms of land use and the spatial organization of economic activity. The level of economic sophistication in the discussion of land use is below that in the discussion of pollution problems.

The basic defect in the land use discussions is the absence of any analysis of market failure. Unlike its discussion of pollution, *CEQ*'s discussion of land use does not include an analysis of external diseconomies or other market failures that it believes create a problem for government policy. *EQ*'s discussion is clearly sympathetic to the notion of government controls on urban decentralization and urban growth. But it has no analysis that would inform readers why it thinks these historical trends are inefficient or inequitable. Instead, the discussion is pejorative and emotional. Suburbanization is explained in terms of the attitudes of the frontier society and of rural values. In fact, the process of urban decentralization has been studied at length by economists and geographers. It pervades the industrialized world and its main causes are widely believed to be rising real incomes and improved transportation. The process is not very different in the United States from other industrialized countries except that incomes are higher and relative prices of land are lower here than elsewhere. There is nothing in the process that suggests market failure. Most of the factors *EQ* points to as causes of urban sprawl are merely market responses to satisfy consumer demands. In the United States, the main reason to believe that urban decentralization has been excessive is that local governments have required excessively low suburban densities by large lot zoning and other land use controls.

Plenty of people can be found to disagree with the above analysis and *CEQ* is no more obligated to accept it than anyone else. But *EQ* merely ignores it and presents no coherent alternative analysis. This is far from an entirely academic issue since federal, state, and local government controls on urban growth and decentralization are now heated public issues. Pollution is only a relatively small part of this matter and there is grave doubt whether *EQ* is the appropriate place for a full discussion of the issues. At least, the analysis they have given does not persuade us.

IV. Conclusions

Our overall evaluation of *EQ* is that it has made a beginning at filling an important public need for a systematic discussion of environmental trends, status, and problems. We believe that the need can be filled much better if future volumes are shorter, better organized, and better focused on pollution problems, with somewhat less attention devoted to peripheral issues. *CEQ* should accord high priority to development of an adequate statistical appendix.

We believe that the economic analysis presented in *EQ* has been of good quality. It has been excessively abbreviated and appears to have had no effect on the discussion of policy developments in other parts of *EQ*. We are concerned at the recent tendency to adopt a defensive stance of restricting economic analysis mainly to demonstrations that the economy can afford existing government programs to abate pollution. Future *EQ*'s should revert to the earlier policy of evaluating benefits as well as costs of environmental programs, and more resources should be devoted to benefit estimation. Our first comprehensive water pollution control legislation was passed almost two decades ago and the analogous air legislation more than one decade ago. We should have a frank evaluation of what has and what has not been accomplished, and which programs have worked better than others. Finally, future volumes of *EQ* should continue the earlier environmental program evaluation that has been absent from recent volumes.

REFERENCES

U.S. Bureau of the Census, *Statistical Abstract of the United States*, Washington, various years.

U.S. Council of Economic Advisors, *Economic Report of the President*, Washington, various years.

U.S. Council on Environmental Quality, *Environmental Quality*, Washington, 1970–74.

Japanese Ministry of Finance, Environment Agency, *Quality of the Environment in Japan*, Tokyo, various years.

Short-Term Interest Rates as Predictors of Inflation

By Eugene F. Fama*

Irving Fisher pointed out that with perfect foresight and a well-functioning capital market, the one-period nominal rate of interest is the equilibrium real return plus the fully anticipated rate of inflation. In a world of uncertainty where foresight is imperfect, the nominal rate of interest can be thought of as the equilibrium expected real return plus the market's assessment of the expected rate of inflation.

The relationships between interest rates and inflation have been tested extensively.[1] In line with Fisher's initial work, the almost universal finding is that there are no relationships between interest rates observed at a point in time and rates of inflation subsequently observed. Although the market does not do well in predicting inflation, the general finding is that there are relationships between current interest rates and past rates of inflation. This is interpreted as evidence in favor of the Fisherian view. Thus Fisher concludes:

> We have found evidence, general and specific, . . . that price changes do, generally and perceptibly, affect the interest rate in the direction indicated by *a priori* theory. But since forethought is imperfect, the effects are smaller than the theory requires and lag behind price movements, in some periods, very greatly. [p. 451]

Fisher's empirical evidence, and that of most other researchers, is in fact inconsistent with a well-functioning or "efficient" market.[2] An efficient market correctly uses all relevant information in setting prices. If the inflation rate is to some extent predictable, and if the one-period equilibrium expected real return does not change in such a way as to exactly offset changes in the expected rate of inflation, then in an efficient market there will be a relationship between the one-period nominal interest rate observed at a point in time and the one-period rate of inflation subsequently observed. If the inflation rate is to some extent predictable and no such relationship exists, the market is inefficient: in setting the nominal interest rate, it overlooks relevant information about future inflation.

This paper is concerned with efficiency in the market for one- to six-month U.S. Treasury Bills. Unlike Fisher and most of the rest of the literature, the results presented here indicate that, at least during the 1953–71 period, there are definite relationships between nominal interest rates and rates of inflation subsequently observed. Moreover, during this period the bill market seems to be efficient in the sense that nominal interest rates summarize all the information about future inflation rates that is in time-series of past inflation rates. Finally, another interesting result is that the substantial variation in nominal bill rates during the 1953–71 period seems to be due entirely to variation in expected inflation rates; in other words, expected real returns on bills seem to be constant during the period.

---

* Professor of finance, Graduate School of Business, University of Chicago. This research was supported by the National Science Foundation. I am grateful to F. Black, N. Gonedes, M. Jensen, A. Laffer, M. Miller, C. Nelson, H. Roberts, and C. Upton for helpful comments, and to J. MacBeth and D. Garren for computational assistance.

[1] For a summary, see Richard Roll.

[2] For a discussion of the theory of efficient capital markets and related empirical work, see the author.

The theory and tests of bill market efficiency are first presented for one-month bills. The results are then extended to bills with longer maturities.

I. Inflation and Efficiency in the Bill Market: Theory

A. *Returns and the Inflation Rate*

The nominal return from the end of month t−1 to the end of month t on a Treasury Bill with one month to maturity at t−1 is

$$(1)\qquad R_t = \frac{v_t - v_{t-1}}{v_{t-1}} = \frac{\$1{,}000 - v_{t-1}}{v_{t-1}}$$

where $v_t = \$1{,}000$ is the price of the bill at t, and v_{t-1} is its price at t−1. Since the bill has one month to maturity at t−1, once v_{t-1} is set, R_t is known and can be interpreted as the one-month nominal rate of interest set in the market at t−1 and realized at t.

Let p_t be the price level at t, that is, p_t is the price of consumption goods in terms of money, so that the purchasing power of a unit of money, the price of money in terms of goods, is $\pi_t = 1/p_t$. The real return from t−1 to t on a one-month bill is then

$$(2)\qquad \tilde{r}_t = (v_t\tilde{\pi}_t - v_{t-1}\pi_{t-1})/v_{t-1}\pi_{t-1}$$

$$(3)\qquad = R_t + \tilde{\Delta}_t + R_t\tilde{\Delta}_t$$

where tildes (˜) are used to denote random variables, and

$$(4)\qquad \tilde{\Delta}_t = (\tilde{\pi}_t - \pi_{t-1})/\pi_{t-1}$$

is the rate of change in purchasing power from t−1 to t. In monthly data, R_t and $\tilde{\Delta}_t$ are close to zero, so that although the equality only holds as an approximation, no harm is done if (3) is reduced to

$$(5)\qquad \tilde{r}_t = R_t + \tilde{\Delta}_t$$

Thus the real return from the end of month t−1 to the end of month t on a Treasury Bill with one month to maturity at t−1 is the nominal return plus the rate of change in purchasing power from t−1 to t.

The fact that $\tilde{r}_t$ is a random variable at t−1 only because $\tilde{\Delta}_t$ is a random variable explains why bills are attractive for studying how well the market uses information about future inflation in setting security prices. It seems reasonable to assume that investors are concerned with real returns on securities. Since all uncertainty in the real return on a one-month bill is uncertainty about the change in the purchasing power of money during the month, one-month bills are the clear choice for studying how well the market absorbs information about inflation one month ahead. For the same reason, *n*-month bills are best for studying *n*-month predictions of inflation.

B. *The General Description of an Efficient Market*

Market efficiency requires that in setting the price of a one-month bill at t−1, the market correctly uses all available information to assess the distribution of $\tilde{\Delta}_t$. Formally, in an efficient market,

$$(6)\qquad f_m(\Delta_t \mid \phi^m_{t-1}) = f(\Delta_t \mid \phi_{t-1})$$

where ϕ_{t-1} is the set of information available at t−1, ϕ^m_{t-1} is the set of information used by the market, $f_m(\Delta_t \mid \phi^m_{t-1})$ is the market assessed density function for $\tilde{\Delta}_t$, and $f(\Delta_t \mid \phi_{t-t})$ is the true density function implied by ϕ_{t-1}.

When the market sets the equilibrium price of a one-month bill at t−1, R_t is also set. Given the relationship among $\tilde{r}_t$, R_t, and $\tilde{\Delta}_t$ of (5), the market's assessed distribution for $\tilde{r}_t$ is implied by R_t and its assessed distribution for $\tilde{\Delta}_t$. If (6) holds, then the market's assessed distribution for $\tilde{r}_t$ is the true distribution

$$(7)\qquad f_m(\tilde{r}_t \mid \phi^m_{t-1}, R_t) = f(\tilde{r}_t \mid \phi_{t-1}, R_t)$$

In short, if the market is efficient, then in setting the nominal price of a one-month bill at t−1, it correctly uses all available information to assess the distribution of $\tilde{\Delta}_t$. In this sense v_{t-1} fully reflects all available information about $\tilde{\Delta}_t$. Since an equilibrium value of v_{t-1} implies an equilibrium value of R_t, the one-month nominal rate of interest set in the market at t−1 likewise fully reflects all available information about $\tilde{\Delta}_t$. Finally, when an efficient market sets R_t, the distribution of the real return $\tilde{r}_t$ that it perceives is the true distribution.

C. *A Simple Model of Market Equilibrium*

The preceding specification of market efficiency is so general that it is not testable. Since we cannot observe $f_m(\Delta_t \mid \phi^m_{t-1})$, we cannot determine whether (6) holds, and so we cannot determine whether the the bill market is efficient. What the model lacks is a more detailed specification of the link between $f_m(\Delta_t \mid \phi^m_{t-1})$ and v_{t-1}; that is, we must specify in more detail how the equilibrium price of a bill at t−1 is related to the market-assessed distribution of $\tilde{\Delta}_t$. This is a common feature of tests of market efficiency. A test of efficiency must be based on a model of equilibrium, and any test is simultaneously a test of efficiency and of the assumed model of equilibrium.

The first assumption of the model of bill market equilibrium is that in their decisions with respect to one-month bills, the primary concern of investors is the distribution of the real return on a bill. A market equilibrium depends visibly on a market-clearing value of the nominal price v_{t-1}, but it is assumed that what causes investors to demand the outstanding supply of bills is the implied "equilibrium distribution" of the real return. Testable propositions about market efficiency then require propositions about the characteristics of the market assessed distribution $f_m(\tilde{r}_t \mid \phi^m_{t-1}, R_t)$ that results from an equilibrium price v_{t-1} at t−1. As is common in tests of market efficiency, we concentrate on the mean of the distribution, and the proposition about $E_m(\tilde{r}_t \mid \phi^m_{t-1}, R_t)$ is that for all t and ϕ^m_{t-1},

$$(8)\qquad E_m(\tilde{r}_t \mid \phi^m_{t-1}, R_t) = E(\tilde{r})$$

Thus the model of bill market equilibrium is the statement that each month the market sets the price of a one-month bill so that it perceives the expected real return on the bill to be $E(\tilde{r})$. In short, the equilibrium expected real return on a one-month bill is assumed to be constant through time.

II. Testable Implications of Market Efficiency When the Equilibrium Expected Real Return Is Constant Through Time

A. *The Real Return*

In an efficient market (7) holds, and (7) implies

$$(9)\qquad E_m(\tilde{r}_t \mid \phi^m_{t-1}, R_t) = E(\tilde{r}_t \mid \phi_{t-1}, R_t)$$

If market equilibrium is characterized by (8), then with (9) we have

$$(10)\qquad E(\tilde{r}_t \mid \phi_{t-1}, R_t) = E(\tilde{r})$$

Thus at any time t−1 the market sets the price of a one-month bill so that its assessment of the expected real return is the constant $E(\tilde{r})$. Since an efficient market correctly uses all available information, $E(\tilde{r})$ is also the true expected real return on the bill.

The general testable implication of this combination of market efficiency with a model of market equilibrium is that there is no way to use ϕ_{t-1}, the set of information available at t−1, or any subset of ϕ_{t-1}, as the basis of a correct assessment of the expected real return on a one-month bill which is other than $E(\tilde{r})$. One subset of ϕ_{t-1} is the time-series of past real returns. If (10) holds,

$$E(\tilde{r}_t \mid r_{t-1}, r_{t-2}, \ldots) = E(\tilde{r}) \tag{11}$$

That is, there is no way to use the time-series of past real returns as the basis of a correct assessment of the expected real return which is other than $E(\tilde{r})$. If (11) holds, the autocorrelations of $\tilde{r}_t$ for all lags are equal to zero, so that sample autocorrelations provide tests of (11).

Sample autocorrelations of $\tilde{r}_t$ are presented later, but it is well to make one point now. The autocorrelations are joint tests of market efficiency and of the model for the equilibrium expected real return. Thus nonzero autocorrelations of $\tilde{r}_t$ are consistent with a world where the equilibrium expected real return is constant and the market is inefficient, but nonzero autocorrelations are also consistent with a world where the market is efficient and equilibrium expected real returns change through time as a function of the sequence of past real returns. Market efficiency in no way rules out such behavior of the equilibrium expected return.

B. *The Nominal Interest Rate as a Predictor of Inflation*

There are tests that distinguish better between the hypothesis that the market is efficient and the hypothesis that the expected real return is constant through time. From (5), the relationship between the market's expectation of the rate of change in purchasing power, the nominal rate of interest, and the market's expectation of the real return is

$$E_m(\tilde{\Delta}_t \mid \phi^m_{t-1}) = E_m(\tilde{r}_t \mid \phi^m_{t-1}, R_t) - R_t \tag{12}$$

If the expected real return is the constant $E(\tilde{r})$, then (12) becomes

$$E_m(\tilde{\Delta}_t \mid \phi^m_{t-1}) = E(\tilde{r}) - R_t \tag{13}$$

If the market is also efficient,

$$E(\tilde{\Delta}_t \mid \phi_{t-1}) = E(\tilde{r}) - R_t \tag{14}$$

Thus a constant expected real return implies that all variation through time in the nominal rate R_t is a direct reflection of variation in the market's assessment of the expected value of $\tilde{\Delta}_t$. If the market is also efficient, then all variation in R_t mirrors variation in the best possible assessment of the expected value of $\tilde{\Delta}_t$. Moreover, once R_t is set at time t−1, the details of ϕ_{t-1}, the information that an efficient market uses to assess the expected value of $\tilde{\Delta}_t$, become irrelevant. The information in ϕ_{t-1} is summarized completely in the value of R_t. In this sense, the nominal rate R_t observed at t−1 is the best possible predictor of the rate of inflation from t−1 to t.

To test these propositions, it is convenient to introduce a new class of models of market equilibrium that includes (8) as a special case. Suppose that at any time t−1 the market always sets the price of a one-month bill so that it perceives the expected real return to be

$$E_m(\tilde{r}_t \mid \phi^m_{t-1}, R_t) = \alpha_0 + \gamma R_t \tag{15}$$

If the market is also efficient, we have

$$E(\tilde{r}_t \mid \phi_{t-1}, R_t) = \alpha_0 + \gamma R_t \tag{16}$$

With (5), (15) and (16) imply that

$$E_m(\tilde{\Delta}_t \mid \phi^m_{t-1}) = \alpha_0 + \alpha_1 R_t, \quad \alpha_1 = \gamma - 1 \tag{17}$$

$$E(\tilde{\Delta}_t \mid \phi_{t-1}) = \alpha_0 + \alpha_1 R_t, \quad \alpha_1 = \gamma - 1 \tag{18}$$

In the new model, γ is the proportion of the change in the nominal rate from one month to the next that reflects a change in the equilibrium expected real return, and $-\alpha_1 = 1 - \gamma$ is the proportion of the change in R_t that reflects a change in the expected value of $\tilde{\Delta}_t$. In the special case where the expected real return is constant through time, $\gamma=0$, $\alpha_1=-1$, and all variation in R_t mirrors variation in $E(\tilde{\Delta}_t|\phi_{t-1})$.

Estimates of α_0 and α_1 in (18) can be obtained by applying least squares to

$$\tilde{\Delta}_t = \alpha_0 + \alpha_1 R_t + \tilde{\epsilon}_t \tag{19}$$

If the coefficient estimates are inconsistent with the hypothesis that

$$\alpha_0 = E(\tilde{r}) \quad \text{and} \quad \alpha_1 = -1 \tag{20}$$

the model of a constant equilibrium expected real return is rejected. The more general interpretation of (15), that is, with unrestricted values of the coefficients, can then be taken as the model for the equilibrium expected real return, and other results from the estimates of (19) can be used to test market efficiency. Thus, like (14), (18) says that in an efficient market R_t summarizes all the information about the expected value of $\tilde{\Delta}_t$ which is in ϕ_{t-1}. For example, given R_t, the sequence of past values of the disturbance $\tilde{\epsilon}_t$ in (19) should be of no additional help in assessing the expected value of $\tilde{\Delta}_t$ which implies that the autocorrelations of the disturbance should be zero for all lags.

The approach is easily generalized to obtain other tests of (14). For example, one item of information available at $t-1$ is Δ_{t-1}. If periods of inflation or deflation tend to persist, then Δ_{t-1} is relevant information for assessing the expected value of $\tilde{\Delta}_t$. If the information in Δ_{t-1} is not correctly used by the market in setting R_t, then the coefficient α_2 in

$$\tilde{\Delta}_t = \alpha_0 + \alpha_1 R_t + \alpha_2 \Delta_{t-1} + \tilde{\epsilon}_t \tag{21}$$

is nonzero. On the other hand, if (14) holds, the market is efficient and the value of R_t set at $t-1$ summarizes all the information available about the expected value of $\tilde{\Delta}_t$, which includes any information in Δ_{t-1} and any information in the past values of $\tilde{\epsilon}_t$. Thus, in this case, $\alpha_2=0$ and the autocorrelations of the disturbance $\tilde{\epsilon}_t$ in (21) are zero for all lags. Moreover, if (14) holds, the expected real return is constant through time, so that the values of α_0 and α_1 in (21) are as in (20). All of these propositions are tested below with least squares estimates of (21).

C. *Reinterpretation of the Proposed Tests*

It is well to recognize that all of the tests of market efficiency are different ways to examine whether in assessing the expected value of $\tilde{\Delta}_t$, the market correctly uses any information in the past values $\Delta_{t-1}, \Delta_{t-2}, \ldots$. The point is obvious with respect to tests based on the coefficient α_2 in (21). The argument is also direct for the autocorrelations of the disturbances $\tilde{\epsilon}_t$ in (19) and (21). The disturbance $\tilde{\epsilon}_t$ in (19) is the deviation of $\tilde{\Delta}_t$ from the market's assessment of its conditional expected value, when $E_m(\tilde{\Delta}_t|\phi^m_{t-1})$ is given by (17). The autocorrelations of $\tilde{\epsilon}_t$ tell us whether the past values of these deviations are used correctly by the market when it assesses the expected value of $\tilde{\Delta}_t$. Nonzero autocorrelations imply that the market is inefficient; one can improve on the market's assessment of the expected value of $\tilde{\Delta}_t$ by making correct use of information in past values of Δ_t. Likewise the disturbance $\tilde{\epsilon}_t$ in (21) is the deviation of $\tilde{\Delta}_t$ from its conditional expected value when the latter is allowed to be a function of Δ_{t-1} as well as of R_t. Finally, if the equilibrium expected real return is constant through time, then the market's assessment of the expected value of $\tilde{\Delta}_t$ is described by (13). From (5) it then follows that

$$\tilde{r}_t - E(\tilde{r}) = \tilde{\Delta}_t + R_t - E(\tilde{r}) \tag{22a}$$

$$= \tilde{\Delta}_t - E_m(\tilde{\Delta}_t \mid \phi^m_{t-1}) \tag{22b}$$

Thus, the deviation of $\tilde{r}_t$ from its expected value is the deviation of $\tilde{\Delta}_t$ from the market's assessment of its expected value, when the latter is described by (13). Tests of market efficiency based on the autocorrelations of $\tilde{r}_t$, like all the other proposed tests, are concerned with whether the market correctly uses any information in the time-series of past values, Δ_{t-1}, Δ_{t-2}, . . . , when it assesses $E_m(\tilde{\Delta}_t | \phi^m_{t-1})$ on which the nominal rate R_t is then based. Any such test must assume some model of market equilibrium, that is, some proposition about the equilibrium expected real return $E_m(\tilde{r}_t | \phi^m_{t-1})$, which in turn implies some proposition about $E_m(\tilde{\Delta}_t | \phi^m_{t-1})$, and this is where the tests differ.

There is, however, no need to apologize for the fact that the tests of market efficiency concentrate on the reaction of the market to information in the time-series of past rates of change in the purchasing power of money. Beginning with the pioneering work of Fisher, researchers in this area have long contended, and the results below substantiate the claim, that past rates of inflation are important information for assessing future rates. Moreover, previous work almost uniformly suggests that the market is inefficient; in assessing expected future rates of inflation, much of the information in past rates is apparently ignored. This conclusion, if true, indicates a serious failing of a free market. The value of a market is in providing accurate signals for resource allocation, which means setting prices that more or less fully reflect available information. If the market ignores the information from so obvious a source as past inflation rates, its effectiveness is seriously questioned. The issue deserves further study.

III. The Data

The one-month nominal rate of interest R_t used in the tests is the return from the end of month t−1 to the end of month t on the Treasury Bill that matures closest to the end of month t. The data are from the quote sheets of Salomon Brothers. In computing R_t from (1), the average of the bid and asked prices at the end of month t−1 is used for the nominal price v_{t-1}. The Bureau of Labor Statistics Consumer Price Index (*CPI*) is used to estimate Δ_t, the rate of change in the purchasing power of money from the end of month t−1 to the end of month t. The use of any index to measure the level of prices of consumption goods can be questioned. There is, however, no need to speculate about the effects of shortcomings of the data on the tests. If the results of the tests seem meaningful, the data are probably adequate.

The tests cover the period from January 1953 through July 1971. Tests for periods prior to 1953 would be meaningless. First, during World War II and up to the Treasury-Federal Reserve Accord of 1951, interest rates on Treasury Bills were pegged by the government. In effect, a rich and obstinate investor saw to it that Treasury Bill rates did not adjust to predictable changes in inflation rates. Second, at the beginning of 1953 there was a substantial upgrading of the *CPI*.[3] The number of items in the Index increased substantially, and monthly sampling of major items became the general rule. For tests of market efficiency based on monthly data, monthly sampling of major items in the *CPI* is critical. Sampling items less frequently than monthly, the general rule prior to 1953, means that some of the price changes for month t show up in the Index in months subsequent to t. Since nominal prices of goods tend to move together, spreading price changes for month t into following months creates spurious positive autocorrelation in monthly changes in the Index. This gives the appearance that there is more information about future inflation rates in past inflation rates than is really the case. Since the spurious component of the information in measured inflation rates is not easily isolated, test of market efficiency on pre-

---

[3] See ch. 10 of the *BLS* reference.

1953 data would be difficult to interpret.

The values of the *CPI* from August 1971 to the present (mid-1974) are also suspect. During this period the Nixon Administration made a series of attempts to fix prices. The controls were effective in creating "shortages" of some important goods (who can forget the gas queues of the winter of 1973–74?), so that for this period there are nontrivial differences between the observed values of the *CPI* and the true costs of goods to consumers. For this reason, the tests concentrate on the "clean" precontrols period January 1953 to July 1971.

IV. Results for One-Month Bills

Table 1 shows sample autocorrelations $\hat{\rho}_\tau$ of Δ_t for lags τ of from one to twelve months. The table also shows sample means and standard deviations of Δ_t, and

$$\sigma(\hat{\rho}_1) = 1/(T-1)^{1/2} \tag{23}$$

where $T-1$ is the number of observations used to compute $\hat{\rho}_1$, and $\sigma(\hat{\rho}_1)$ is the approximate standard error of $\hat{\rho}_1$ under the hypothesis that the true autocorrelation is zero. Table 2 shows sample autocorrelations and other statistics for the real return $\tilde{r}_t$. Although, for simplicity, the development of the theory is in terms of the approximation given by (5), the exact expression (3) is used to compute $\tilde{r}_t$ in the empirical work.

Table 3 shows summary statistics for the estimated version of (19). In addition to the least squares regression coefficient estimates a_0 and a_1, the table shows the sample standard errors of the estimates $s(a_0)$ and $s(a_1)$; the coefficient of determination, adjusted for degrees of freedom; $s(e)$, the standard deviation of the residuals; and the first three residual autocorrelations, $\hat{\rho}_1(e)$, $\hat{\rho}_2(e)$, and $\hat{\rho}_3(e)$. Table 4 shows similar summary statistics for the estimated version of (21).

A. *The Information in Past Inflation Rates*

The market efficiency hypothesis to be tested is that the one-month nominal interest rate R_t set in the market at the end of month t−1 is based on correct utilization of all the information about the expected value of $\tilde{\Delta}_t$, which is in the time-series of past values Δ_{t-1}, Δ_{t-2}, ... The hypothesis is only meaningful, however, if past rates of change in purchasing power do indeed have information about the expected future rate of change. The predominance

TABLE 1—AUTOCORRELATIONS OF Δ_t: ONE-MONTH INTERVALS

	1/53–7/71	1/53–2/59	3/59–7/64	8/64–7/71
$\hat{\rho}_1$	.36	.21	−.09	.35
$\hat{\rho}_2$	.37	.28	−.09	.34
$\hat{\rho}_3$	.27	.10	−.25	.26
$\hat{\rho}_4$	.30	.16	−.05	.23
$\hat{\rho}_5$	.29	.01	.03	.33
$\hat{\rho}_6$	.29	−.01	.09	.30
$\hat{\rho}_7$	.25	.05	−.06	.18
$\hat{\rho}_8$	.34	.18	−.20	.37
$\hat{\rho}_9$	.36	.21	.13	.24
$\hat{\rho}_{10}$	.34	.20	.04	.21
$\hat{\rho}_{11}$	.27	.09	−.09	.18
$\hat{\rho}_{12}$	.37	.18	.17	.30
$\sigma(\hat{\rho}_1)$	.07	.12	.13	.11
$\bar{\Delta}$	−.00188	−.00111	−.00108	−.00321
$s(\Delta)$	.00234	.00258	.00169	.00195
$T-1$	222	73	64	83

TABLE 2—AUTOCORRELATIONS OF r_t: ONE-MONTH BILLS

	1/53–7/71	1/53–2/59	3/59–7/64	8/64–7/71
$\hat{\rho}_1$	.09	.11	−.04	.10
$\hat{\rho}_2$	.13	.17	.01	.08
$\hat{\rho}_3$	−.02	−.02	−.20	−.01
$\hat{\rho}_4$	−.01	.01	−.06	−.10
$\hat{\rho}_5$	−.02	−.14	.00	.08
$\hat{\rho}_6$	−.02	−.18	.07	.07
$\hat{\rho}_7$	−.07	−.09	−.09	−.15
$\hat{\rho}_8$	.04	.05	−.23	.17
$\hat{\rho}_9$	.11	.11	.09	.04
$\hat{\rho}_{10}$	.10	.12	.07	−.02
$\hat{\rho}_{11}$	.03	.03	−.10	−.07
$\hat{\rho}_{12}$	.19	.16	.19	.15
$\sigma(\hat{\rho}_1)$	.07	.12	.13	.11
$\bar{r}$	.00074	.00038	.00111	.00075
$s(r)$	.00197	.00240	.00172	.00168
$T-1$	222	73	64	83

TABLE 3—REGRESSION TESTS ON ONE-MONTH BILLS
$\Delta_t = a_0 + a_1 R_t + e_t$

Period	a_0	a_1	$s(a_0)$	$s(a_1)$	Coefficient of Determination	$s(e)$	$\hat{\rho}_1(e)$	$\hat{\rho}_2(e)$	$\hat{\rho}_3(e)$
1/53–7/71	.00070	−.98	.00030	.10	.29	.00196	.09	.13	−.02
1/53–2/59	.00116	−1.49	.00069	.42	.14	.00240	.09	.15	−.05
3/59–7/64	−.00038	−.33	.00095	.42	−.01	.00168	−.09	−.08	−.26
8/64–7/71	.00118	−1.10	.00083	.20	.26	.00167	.09	.06	−.02

of large estimated autocorrelations of Δ_t in Table 1 indicates that this is the case.

In fact, especially for the longer periods 1/53–7/71 and 8/64–7/71, the sample autocorrelations of Δ_t for different lags are similar in size with individual estimates in the neighborhood of .30. This finding is discussed later when the behavior through time of Δ_t is studied in more detail.

B. *Market Efficiency*

Given that the equilibrium expected real return is constant through time, the market efficiency hypothesis says that the autocorrelations of the real return $\tilde{r}_t$ are zero for all lags. The sample autocorrelations of r_t in Table 2 are close to zero. Recall from (5) that the real return $\tilde{r}_t$ is approximately the rate of change in purchasing power Δ_t plus the nominal interest rate R_t. The evidence from the sample autocorrelations of Δ_t and r_t in Tables 1 and 2 is that adding R_t to Δ_t brings the substantial autocorrelations of Δ_t down to values close to zero. This is consistent with the hypothesis that R_t, the nominal rate set at $t-1$, summarizes completely the information about the expected value of $\tilde{\Delta}_t$ which is in the time-series of past values, $\Delta_{t-1}, \Delta_{t-2}, \ldots$.

Tables 3 and 4 give further support to the market efficiency hypothesis. When applied to (21), the hypothesis says that α_2, the coefficient of Δ_{t-1}, is zero, and the autocorrelations of the disturbance $\tilde{\epsilon}_t$ are likewise zero for all lags. The residual autocorrelations in Table 4 are close to zero. The values of a_2, the sample estimates of α_2 in (21), are also small and always less than two standard errors from zero. When applied to (19), the market efficiency hypothesis is again that the autocorrelations of the disturbance $\tilde{\epsilon}_t$ should be zero. The residual autocorrelations in Table 3 are close to zero. Moreover, comparing the results for the estimated versions of (19) and (21) in Tables 3 and 4 shows that dropping Δ_{t-1} from the model has almost no effect on the coefficients of determination, which is consistent with the implication of market efficiency that the value of R_t set at time $t-1$ summarizes any information in Δ_{t-1} about the expected value of $\tilde{\Delta}_t$.

Closer inspection of the tables seems to

TABLE 4—REGRESSION TESTS ON ONE-MONTH BILLS
$\Delta_t = a_0 + a_1 R_t + a_2 \Delta_{t-1} + e_t$

Period	a_0	a_1	a_2	$s(a_0)$	$s(a_1)$	$s(a_2)$	Coefficient of Determination	$s(e)$	$\hat{\rho}_1(e)$	$\hat{\rho}_2(e)$	$\hat{\rho}_3(e)$
1/53–7/71	.00059	−.87	.11	.00030	.12	.07	.30	.00195	−.05	.13	−.04
1/53–2/59	.00108	−1.40	.11	.00069	.44	.11	.14	.00238	−.09	.17	−.07
3/59–7/64	−.00054	−.30	−.08	.00097	.42	.13	−.02	.00170	−.01	−.11	−.25
8/64–7/71	.00073	−.89	.14	.00084	.24	.11	.24	.00164	−.04	.05	−.01

provide slight evidence against market efficiency. Except for the 3/59–7/64 period, the first-order sample autocorrelations of r_t, though small, are nevertheless all positive. The estimated regression coefficients a_2 of Δ_{t-1} in Table 4 are likewise small but generally positive, as are the first-order residual autocorrelations in Table 3. It is well to note, however, that even after the upgrading of the *CPI* in 1953, there are some items whose prices are sampled less frequently than monthly; and items that are sampled monthly are not sampled at the same time during the month. Again, since prices of goods tend to move together, these quirks of the sampling process induce spurious positive autocorrelation in measured rates of change in purchasing power. Since an efficient market does not react to "information" that is recognizably spurious, the small apparent discrepancies from efficiency provide more "reasonable" evidence in favor of the efficiency hypothesis than if the data suggested that the hypothesis does perfectly well.

C. *The Expected Real Return*

The evidence is also consistent with the hypothesis that the expected real return on a one-month bill is constant during the 1953–71 period. First, the sample autocorrelations of the real return $\tilde{r}_t$ are joint tests of the hypotheses that the market is efficient and that the expected real return is constant through time. Since the sample autocorrelations of r_t in Table 2 are close to zero, the evidence is consistent with a world where both hypotheses are valid.

The regression coefficient estimates for (19) and (21) in Tables 3 and 4 are, however, more direct evidence on the hypothesis that the expected value of $\tilde{r}_t$ is constant. The hypothesis implies that in (19) and (21), the intercept α_0 is the constant expected real return $E(\tilde{r})$ and the coefficient α_1 of R_t is -1.0. The coefficient estimates a_1 of α_1 in (19) and (21) are always well within two standard errors of -1.0. And statistical considerations aside, the estimate $a_1=-.98$ for (19) for the overall period 1/53–7/71 is impressively close to -1.0. Given estimates a_1 of α_1 in (19) and (21) that are close to -1.0, and given the earlier observation that the estimates a_2 of α_2 in (21) are close to zero, equation (5) and the least squares formulas guarantee that the intercept estimates a_0 for (19) and (21) in Tables 3 and 4 are close to the sample means of the real return in Table 2.

Finally, the sample autocorrelations of r_t in Table 2 and the regression coefficient estimates a_0 and a_1 in Tables 3 and 4 are consistent with the world of equation (13) where the equilibrium expected real return is constant and all variation through time in the nominal interest rate R_t mirrors variation in the market's assessment of the expected value of $\tilde{\Delta}_t$. There is, however, another interesting way to check this conclusion. From the discussion of (22) it follows that the standard deviation of the real return $\tilde{r}_t$ is the standard deviation of the disturbance $\tilde{\epsilon}_t$ in (19) when the coefficients α_0 and α_1 in (19) are constrained to have the values $\alpha_0=E(r)$ and $\alpha_1=-1.0$ that are appropriate under the hypothesis that the expected real return is constant through time. If this hypothesis is incorrect, letting the data choose values of α_0 and α_1, as in Table 3, should produce lower estimates of the disturbance variance than when the values of the coefficients are constrained. But the results indicate that, especially for the longer periods, not only are the values of $s(r)$ in Table 2 almost identical to the values of $s(e)$ in Table 3, but the sample autocorrelations of r_t and e_t are almost identical. In short, the hypothesis that the expected real return is constant fits the data so well that the residuals from the estimated version of (19) are more or less identical to the deviations of $\tilde{r}_t$ from its sample mean.

V. The Behavior of $\tilde{\Delta}_t$

The results allow some interesting insights into the behavior through time of $\tilde{\Delta}_t$. The rate of change in purchasing power can always be written as

$$\tilde{\Delta}_t = E(\tilde{\Delta}_t \mid \phi_{t-1}) + \tilde{\epsilon}_t \tag{24}$$

Since the evidence is consistent with the hypothesis that the expected real return is constant through time, we can substitute (14) into (24) to get

$$\tilde{\Delta}_t = E(\tilde{r}) - R_t + \tilde{\epsilon}_t \tag{25}$$

The conclusion drawn from the residual autocorrelations in Table 3 and the sample autocorrelations of $\tilde{r}_t$ in Table 2 is that the disturbance $\tilde{\epsilon}_t$ in (25) is uncorrelated through time. The time-series of past values of $\tilde{\epsilon}_t$ is no real help in predicting the next value. Quite the opposite sort of behavior characterizes the expected value of $\tilde{\Delta}_t$ in (24). Since, as stated in (25), variation in R_t through time mirrors variation in the expected value of $\tilde{\Delta}_t$, the time-series properties of R_t are the time-series properties of $E(\tilde{\Delta}_t \mid \phi_{t-1})$. For the 1/53–7/71 period, the first four sample autocorrelations of R_t are all in excess of .93, and only one of the first twenty-four is less than .9. Sample autocorrelations close to 1.0 are consistent with the representation of R_t as a random walk. Thus in contrast with the evidence for the disturbance $\tilde{\epsilon}_t$ in (24), the autocorrelations of R_t indicate that there is much persistence through time in the level of R_t and thus in the level of $E(\tilde{\Delta}_t \mid \phi_{t-1})$. The time-series of past values of R_t has substantial information about future values.

This discussion helps explain the behavior of the sample autocorrelations of $\tilde{\Delta}_t$ in Table 1. As stated in (24), $\tilde{\Delta}_t$ has two components. One component of $\tilde{\Delta}_t$, its expected value, behaves like a random walk. The other component of $\tilde{\Delta}_t$, the disturbance $\tilde{\epsilon}_t$, is essentially random noise. The autocorrelations of its expected value cause the autocorrelations of $\tilde{\Delta}_t$ to likewise have approximately the same magnitude for different lags. The uncorrelated disturbance $\tilde{\epsilon}_t$, however, causes the autocorrelations of $\tilde{\Delta}_t$, unlike those of R_t, to be far below 1.0.

The sample autocorrelations of R_t suggest that the expected value of $\tilde{\Delta}_t$ behaves through time much like a random walk. The sample autocorrelations of the month-to-month changes in R_t, shown in Table 5, suggest, however, that we can improve on this description of the behavior of $E(\tilde{\Delta}_t \mid \phi_{t-1})$. For example, the first-order autocorrelations of $R_t - R_{t-1}$ are consistently negative. From the first-order autocorrelations for the longer periods, the change in R_t might reasonably be represented as

$$\tilde{R}_{t+1} - R_t = -.25(R_t - R_{t-1}) + \tilde{\eta}_t \tag{26}$$

Thus the process that generates the nominal rate is no longer just a random walk. The process is slightly regressive so that on average the change in the expected inflation rate from one month to the next reverses itself by about 25 percent.

TABLE 5—AUTOCORRELATIONS OF $R_t - R_{t-1}$

	1/53–7/71	1/53–2/59	3/59–7/64	8/64–7/71
$\hat{\rho}_1$	−.25	−.14	−.41	−.18
$\hat{\rho}_2$	.06	.05	.07	.06
$\hat{\rho}_3$	.01	.07	−.03	.00
$\hat{\rho}_4$	.15	.23	.08	.18
$\hat{\rho}_5$	−.03	−.04	.07	−.13
$\hat{\rho}_6$	−.06	.01	−.12	−.01
$\hat{\rho}_7$	−.13	−.35	−.11	−.05
$\hat{\rho}_8$	.10	.17	.13	.02
$\hat{\rho}_9$	.06	−.03	−.06	.18
$\hat{\rho}_{10}$	−.24	−.26	−.07	−.42
$\hat{\rho}_{11}$	−.05	−.16	−.10	.08
$\hat{\rho}_{12}$	.09	.13	.06	.04
$\sigma(\hat{\rho}_1)$	.07	.12	.13	.11
$\overline{dR}$	.00001	.00000	.00001	.00001
$s(dR)$	.00032	.00028	.00035	.00033
$T-1$	221	72	63	82

VI. Results for Bills with Longer Maturities

The presentation of theory and tests of bill market efficiency has concentrated so far on one-month bills and one-month rates of change in the purchasing power of money. As far as the theory is concerned, the interval of time over which the variables are measured is arbitrary. In testing the theory, the fact that the *CPI* is only reported monthly limits us to tests based on intervals that cover an integral number of months. Tests are presented now for one- to six-month intervals. Thus, in these tests the interval from $t-1$ to t is one, or two, . . . , or six months; R_t is the sure one-, or two-, . . . , or six-month nominal rate of interest from $t-1$ to t on a bill with one, or two, . . . , or six months to maturity at $t-1$; and the real return $\tilde{r}_t$ and the rate of change in the purchasing power of money $\tilde{\Delta}_t$ are likewise measured for nonoverlapping one- to six-month intervals.

Since the theory and tests are the same for bills of all maturities, the market efficiency hypothesis is that in setting the nominal rate R_t at time $t-1$, the market correctly uses any information about the expected value of $\tilde{\Delta}_t$ which is in the time-series of past values $\Delta_{t-1}, \Delta_{t-2}, \ldots$. The model of market equilibrium on which the tests are based is the assumption that the expected real returns on bills with one to six months to maturity are constant through time. The tests of these propositions are in Tables 6 to 9, and the tests are the same as those for one-month bills in Tables 1 to 4. Results for the one- to three-month versions of the variables are shown for the 1/53–7/71 and 3/59–7/71 periods. Since the data for four- to six-month bills are only available beginning in March 1959, results for the four- to six-month versions of the variables are only shown for the 3/59–7/71 period.

Implicit in the tests of market efficiency is the assumption that past rates of change in purchasing power have information about expected future rates of change. The autocorrelations of Δ_t in Table 6 support this assumption. The autocorrelations are large for all six intervals used to measure Δ_t. But consistent with the hypotheses that the market is efficient and that the equilibrium expected real returns on bills with different maturities are constant through time, the autocorrelations of the real returns shown in Table 7 are close to zero. Remember from (5) that the n-month real return on an n-month bill is approximately the n-month rate of change in purchasing power plus the n-month nominal return on the bill. Thus the evidence from the autocorrelations of Δ_t and r_t in Tables 6 and 7 is that when R_t is added to Δ_t, the substantial autocorrelations of Δ_t drop to values close to zero. This is consistent with a world where R_t, the n-month nominal rate set at $t-1$, summarizes all the information about the expected value of the rate of change in purchasing power over the n months from $t-1$ to t which is in the time-series of past rates of changes in purchasing power.

The model gets further support from the regression tests in Table 8. Consistent with the hypothesis that expected real returns are constant through time, the estimates a_1 of α_1 in (19) in Table 8 are all impressively close to -1.0. Consistent with the hypothesis that the market is efficient, the residual autocorrelations in Table 8 are close to zero for bills of all maturities.

The only hint of evidence against the model is in the estimates of (21) for five- and six-month bills in Table 9. As predicted by the model, the values of a_1 and a_2 for one- to four-month bills are close to -1.0 and 0.0, and the residual autocorrelations are close to 0.0. For the five- and six-month bills, however, the values of a_1 are rather far from -1.0 and the values of a_2

TABLE 6—AUTOCORRELATIONS OF Δ_t: ONE- TO-SIX-MONTH INTERVALS

		1/53–7/71 Interval				3/59–7/71 Interval			
	1	2	3	1	2	3	4	5	6
$\hat{\beta}_1$	.36	.50	.53	.40	.55	.58	.67	.84	.86
$\hat{\beta}_2$	.37	.39	.57	.39	.50	.74	.72	.78	.83
$\hat{\beta}_3$	.27	.43	.59	.32	.66	.64	.71	.74	.74
$\hat{\beta}_4$	.30	.45	.54	.36	.57	.70	.71	.73	.81
$\hat{\beta}_5$	.29	.52	.48	.43	.58	.66	.63	.76	.90
$\hat{\beta}_6$	.29	.41	.38	.44	.56	.65	.76	.77	1.03[a]
$\hat{\beta}_7$	.25	.40	.39	.34	.53	.65	.61	.89	.98
$\hat{\beta}_8$	.34	.32	.27	.40	.60	.58	.73	.83	.95
$\hat{\beta}_9$	.36	.36	.32	.44	.55	.73	.70	.94	.45
$\hat{\beta}_{10}$	.34	.30	.08	.40	.49	.42	.65	.79	.32
$\hat{\beta}_{11}$	.27	.28	.35	.34	.54	.84	.54	.14	−.07
$\hat{\beta}_{12}$	.37	.28	.29	.47	.56	.55	.82	.11	.23
$\sigma(\hat{\beta}_1)$	.07	.10	.12	.08	.12	.14	.17	.19	.21
$\bar{\Delta}$	−.00188	−.00368	−.00550	−.00228	−.00445	−.00656	−.00881	−.01105	−.01319
$s(\Delta)$	.00234	.00386	.00521	.00211	.00348	.00485	.00628	.00735	.00857
$T-1$	222	110	73	148	73	49	36	29	24

[a] The sample autocorrelations are estimated as linear regression coefficients. Thus the estimates can be greater than 1.0.

are rather far from 0.0. In conducting so many different tests for so many different bills, however, some results are likely to turn out badly even though the model is a valid approximation to the world. This argument gains force from the fact that the autocorrelations of the real returns in Table 7 and the estimates of (19) in Table 8 do not produce evidence for five- and six-month bills that contradicts the model.

VII. Interest Rates as Predictors of Inflation: Comparisons with the Results of Others

In a world where equilibrium expected real returns on bills are constant through

TABLE 7—SAMPLE AUTOCORRELATIONS OF r_t: ONE- TO SIX-MONTH BILLS

		1/53–7/71 Bill				3/59–7/71 Bill			
	1	2	3	1	2	3	4	5	6
$\hat{\beta}_1$	.09	.15	.00	.05	.03	−.16	−.17	.02	.07
$\hat{\beta}_2$	.13	−.09	.02	.05	−.15	.16	−.06	−.13	.07
$\hat{\beta}_3$	−.02	−.03	.08	−.08	.18	−.14	.20	−.03	−.05
$\hat{\beta}_4$	−.01	.01	.26	−.07	−.06	.25	.14	.03	.26
$\hat{\beta}_5$	−.02	.18	.16	.06	.00	.11	−.08	.15	.11
$\hat{\beta}_6$	−.02	.10	−.09	.10	.10	.04	.30	−.19	.43
$\hat{\beta}_7$	−.07	.15	.06	−.10	.07	.06	−.22	.33	−.04
$\hat{\beta}_8$	.04	−.01	−.01	.00	.14	.02	.17	−.02	.49
$\hat{\beta}_9$	.11	.06	.08	.09	.08	.18	.16	.25	−.60
$\hat{\beta}_{10}$	.10	.00	−.32	.05	−.07	−.33	−.09	.04	.27
$\hat{\beta}_{11}$	.03	.04	.11	−.04	.08	.36	−.02	−.69	.32
$\hat{\beta}_{12}$	.19	.09	.19	.20	.20	.10	.32	.13	.07
$\sigma(\hat{\beta}_1)$	.07	.10	.12	.08	.12	.14	.17	.19	.21
$\bar{r}$	.00074	.00185	.00306	.00090	.00224	.00373	.00514	.00706	.00882
$s(r)$	.00197	.00292	.00371	.00169	.00236	.00307	.00379	.00375	.00444
$T-1$	222	110	73	148	73	49	36	29	24

TABLE 8—REGRESSION TESTS ON ONE- TO SIX-MONTH BILLS
$\Delta_t = a_0 + a_1 R_t + e_t$

Period	Bill	a_0	a_1	$s(a_0)$	$s(a_1)$	Coefficient of Determination	$s(e)$	$\hat{\rho}_1(e)$	$\hat{\rho}_2(e)$	$\hat{\rho}_3(e)$
1/53–7/71	1	.00070	− .98	.00030	.10	.29	.00196	.09	.13	−.02
	2	.00161	− .96	.00066	.11	.42	.00296	.15	−.08	−.03
	3	.00228	− .92	.00105	.11	.48	.00380	.00	.03	.10
3/59–7/71	1	.00120	−1.09	.00041	.12	.36	.00169	.04	.05	−.08
	2	.00269	−1.08	.00086	.12	.52	.00245	.02	−.16	.14
	3	.00397	−1.03	.00145	.13	.55	.00330	−.16	.12	−.16
	4	.00543	−1.03	.00216	.14	.58	.00413	−.18	−.10	.14
	5	.00635	− .97	.00236	.12	.68	.00416	.01	−.10	−.02
	6	.00879	−1.01	.00344	.14	.65	.00505	.01	−.01	−.11

time, then, aside from the additive constant $E(\tilde{r})$ in (13), the nominal rate R_t set at time t−1 is in effect the market's prediction of the rate of change in purchasing power from t−1 to t. The coefficients of determination in Table 8 indicate that variation through time in these predictions accounts for 30 percent of the variance of subsequently observed values of Δ_t in the case of one-month bills, and the proportion of the sample variance of Δ_t accounted for by R_t increases to about 65 percent for five- and six-month bills. Thus, nominal interest rates observed at t−1 contain nontrivial information about the rate of change in purchasing power from t−1 to t. Moreover, the evidence on market efficiency suggests that the market's prediction of $\tilde{\Delta}_t$ is the best that can be made on the basis of information available at time t−1; or, more precisely, it is the best that can be done on the basis of information in past rates of change in purchasing power.

As noted earlier, the results reported here differ substantially from those of the rest of the literature on interest rates and inflation. In line with the early work of Fisher, the almost universal finding in other studies is that the market does not perform efficiently in predicting inflation. But the earlier studies, including, of course, Fisher's, are based primarily on pre-1953 data, and the negative results on market efficiency may to a large extent just reflect poor commodity price data. By

TABLE 9—REGRESSION TESTS ON ONE- TO SIX-MONTH BILLS
$\Delta_t = a_0 + a_1 R_t + a_2 \Delta_{t-1} + e_t$

Period	Bill	a_0	a_1	a_2	$s(a_0)$	$s(a_1)$	$s(a_2)$	Coefficient of Determination	$s(e)$	$\hat{\rho}_1(e)$	$\hat{\rho}_2(e)$	$\hat{\rho}_3(e)$
1/53–7/71	1	.00059	− .87	.11	.00030	.12	.07	.30	.00195	−.05	.13	−.04
	2	.00115	− .78	.17	.00064	.13	.09	.44	.00280	.03	−.06	.02
	3	.00173	− .79	.11	.00107	.15	.12	.48	.00372	−.06	.07	.05
3/59–7/71	1	.00109	−1.01	.07	.00042	.14	.08	.35	.00169	−.03	.05	−.07
	2	.00252	−1.02	.05	.00094	.18	.12	.51	.00248	−.02	−.16	.15
	3	.00390	−1.06	−.04	.00169	.23	.17	.53	.00334	−.10	.11	−.17
	4	.00520	− .97	.07	.00261	.26	.20	.57	.00423	−.23	−.06	.12
	5	.00359	− .57	.40	.00301	.27	.23	.71	.00404	−.13	−.08	−.02
	6	.00263	− .39	.58	.00406	.28	.23	.72	.00461	−.29	.18	−.32

the same token, the success of the tests reported here is probably to a nonnegligible extent a consequence of the availability of good data beginning in 1953.

Poor commodity price data also probably explain why the empirical literature is replete with evidence in support of the so-called Gibson Paradox—the proposition that there is a positive relationship between the nominal interest rate and the level of commodity prices, rather than the relationship between the interest rate and the rate of change in prices posited by Fisher.[4] With a poor price index, the Fisherian relationship between the nominal interest rate and the true inflation rate can be obscured by noise and by spurious autocorrelation in measured inflation rates. But over long periods of time—and the Gibson Paradox is usually posited as a long-run phenomenon—even a poor index picks up general movements in prices. Thus if inflations and deflations tend to persist (an implication of the evidence presented here that $E(\tilde{\Delta}_t|\phi_{t-1})$ is close to a random walk), there may well appear to be a relationship between the level of interest rates and the measured level of prices, which merely reflects the more fundamental Fisherian relationship between the interest rate and the rate of change of prices that is obscured by poor data. In this study, which is based on the relatively clean data of the 1953–71 period, the Fisherian relationship shows up clearly.

VIII. Conclusions

The two major conclusions of the paper are as follows. First, during the 1953–71 period, the bond market seems to be efficient in the sense that in setting one- to six-month nominal rates of interest, the market correctly uses all the information about future inflation rates that is in time-series of past inflation rates. Second, one cannot reject the hypothesis that equilibrium expected real returns on one- to six-month bills are constant during the period. When combined with the conclusion that the market is efficient, this means that one also cannot reject the hypothesis that all variation through time in one- to six-month nominal rates of interest mirrors variation in correctly assessed one- to six-month expected rates of change in purchasing power.

REFERENCES

E. F. Fama, "Efficient Capital Markets: A Review of Theory and Empirical Work," *J. Finance*, May 1970, *25*, 383–417.

I. Fisher, *The Theory of Interest*, New York 1930, reprinted A. M. Kelley, 1965.

R. Roll, "Interest Rates on Monetary Assets and Commodity Price Index Changes," *J. Finance*, May 1972, *27*, 251–77.

T. J. Sargent, "Interest Rates and Prices in the Long Run: A Study of the Gibson Paradox," *J. Money, Credit, Banking*, Feb. 1973, *5*, 385–449.

U.S. Bureau of Labor Statistics, *Handbook of Methods for Surveys and Studies*, Bull. 1458, Washington 1971.

Salomon Brothers, "United States Treasury Securities," New York, issued daily.

---

[4] For a discussion of the Gibson Paradox and a review of previous evidence, see Roll. A more recent study is Thomas Sargent.

The Theory of "Screening," Education, and the Distribution of Income

By JOSEPH E. STIGLITZ*

One of the most important kinds of information concerns the *qualities* of a factor or a commodity. We know that there are important differences among individuals, among bonds, among equities, among brands of automobiles. The identification of these qualities we call *screening*, and devices that sort our *commodities* (individuals) according to their qualities we call *screening devices* (for example, egg sorters).

This paper focuses on the labelling of individuals, on the economic costs and benefits of labelling, the institutions that provide it, and the determination of the equilibrium amount of screening under various institutional arrangements.

Economists have traditionally argued that because of the problem of appropriability in a market context, too few resources will be allocated to obtaining "information." This is not the case with the information provided by screening processes: individuals who can be labelled as "more productive" are able thereby to obtain a higher wage, partly, however, at the expense of others. Thus, by its very nature, screening information has important effects on the distribution of income.

The basic argument of this paper is that economies with imperfect information with respect to qualities of individuals differ in fundamental ways from economies with perfect information. There may be, for instance, multiple equilibria in which one of the equilibria is Pareto inferior to another; the Pareto inferior equilibrium may involve either too much or too little screening, or it may entail the wrong kind of screening. On the other hand, there may be situations where there exists no equilibrium.

The paper is divided into two parts. In

---

* Professor of economics, Cowles Foundation, Yale University, and Visiting Fellow, St. Catherine's College, Oxford. The research described in this paper was conducted in part while I was a Research Fellow at the Institute for Development Studies, University of Nairobi, 1969–71 under a grant from the Rockefeller Foundation. Financial support from the Ford and National Science Foundation is also gratefully acknowledged. I am indebted to Gary Fields, Michael Rothschild, Michael Spence, and to participants at seminars at Yale, Pennsylvania, Chicago, Queens, Wesleyan, and Princeton at which earlier versions of this paper were presented, for helpful conversations and comments.

The idea that education serves as a screening device and that as a result the allocation of resources to education may not be optimal is, of course, an old one. (See, for instance, Hull and Peters (1969) and Young (1958).) More recently, Thurow, Fields, Akerlof (1973), Spence, and Arrow (1973) have discussed education as a screening device. The first two papers assume a disequilibrium in the labor market (i.e., wages of any group of individuals need not equal the mean marginal product of the group); none of the papers, with the exception of Fields, appears to contain a completely articulated theory of the equilibrium of the system (the "supply" of education), and without a theory of the determination of the screening mechanism, it is difficult to make welfare economic evaluations of the system. For a more extensive discussion of this point, see Rothschild and Stiglitz (1973a). Several of the results are closely related to those obtained independently by Akerlof (1973), Spence, and Arrow (1973). As we show below, the presumption that these papers attempt to establish, that there is too much screening, is not necessarily valid. Various aspects of the theory of screening have recently been the subject of extensive discussion in other areas besides those of the capital market and education referred to earlier: in insurance markets (Rothschild and Stiglitz (1973a)), in labor markets (Salop and Salop (1972)), in discrimination (Arrow (1972), Phelps, Stiglitz (1973, 1974)), and in product markets (Salop (1973)). See also Akerlof's (1970) seminal work on the theory of lemons.

Section I, I develop, partly by means of a number of examples, the central aspects of the theory of screening. Section II is devoted to an analysis of the implications of screening for the allocation of resources to education.

I. The Theory of Screening

A. *The Benefits and Costs of Screening: Private Returns*

We begin with the simplest possible example involving screening. All our later examples (and the examples of George Akerlof (1973), Kenneth Arrow (1973), Michael Spence, J. K. Salop and S. C. Salop, and Michael Rothschild and the author (1973a)) can be thought of as elaborations—on the screening mechanism, the production technology, etc.—of this example.

Consider a population in which individuals can be described (at least for economic purposes) by a single characteristic, which we denote by θ, and which is proportional to the individual's productivity p:

$$p = m\theta$$

(That is, an individual of type θ_2 can do in an hour what a worker of type θ_1 can do in θ_1/θ_2 hours.) The variable p can be interpreted as the individual's marginal product. We choose our units so that $m=1$. The fraction of the population that is of type θ is given by $h(\theta)$.

Assume that the individual knows his ability but the market does not, and in the absence of any information treats all individuals identically. Firms are risk neutral, and act competitively. Assume moreover that the individual is assigned to an assembly line, and on that assembly line it is impossible to tell the productivity of any single individual without prohibitively costly examination. The output per man of the assembly line is proportional to the average value of θ for those working on the assembly line, and there are no other factors of production.

Under these assumptions, a worker will receive a wage equal to the mean value of those with whom he is grouped. If individuals with higher θ can be identified, they will receive a higher wage. They thus have an economic incentive to be identified.

Consider a case where there are only two groups, denoted by θ_1 and θ_2, $\theta_1 > \theta_2$, and which we refer to as the more able and less able, respectively. Assume there is a screening process which screens perfectly[1] and which costs c per individual screened, where

$$(1) \qquad \theta_1 - \theta_2 > c > \theta_1 - \bar{\theta}$$

$$(2) \qquad \bar{\theta} = \theta_1 h(\theta_1) + \theta_2(1 - h(\theta_1))$$

$$= \text{average value of } \theta$$

First we consider a case where the supply of labor by each individual is inelastic, so that with perfect knowledge, the first group would receive an income of θ_1 and the second an income of θ_2. These are best thought to be lifetime incomes, i.e., present discounted values of wage streams.

We now establish that there are two equilibria:

(a) *The no-screening equilibrium.* Since no differentiation is made among individuals, they will all receive the same income, equal to the mean productivity of the population, $\bar{\theta}$. To see that this is an equilibrium observe that it does not pay any individual, in particular, it does not pay the more able individual, to be screened. For with screening, he would obtain a gross income of θ_1, from which we must subtract the cost of screening to obtain net income, $\theta_1 - c$, and by (1), this is less than the income he would have received in the absence of screening, $\bar{\theta}$.

---

[1] Implicitly, we assume that the technology of screening is such that if less than c is spent, there is no screening, i.e., labels are assigned randomly.

(b) *The full-screening equilibrium.* The individuals of type θ_1 receive a gross income of θ_1, a net income of $\theta_1 - c$ (after paying for screening costs); individuals of type θ_2 receive an income of θ_2. Since these individuals know that they are the less able, they do not pay for any screening. Clearly, it pays individuals of type 1 to pay for screening: By our assumptions, all individuals who are not screened are "lumped" together and receive the same wage, so an individual of type 1 who is not screened would have received an income of θ_2, which by (1) is less than his net income with screening.

This simple example illustrates four propositions concerning economies with screening:

1) *There may be multiple equilibria.*[2]

2) *Some of the equilibria are unambiguously Pareto inferior to other equilibria.* Note that in the full-screening equilibrium, both groups have lower net incomes than in the no-screening equilibrium: the first group has an income of $\theta_1 - c$, which by (1) is less than $\bar{\theta}$; the second group an income of θ_2 which is obviously less than $\bar{\theta}$.

3) In both equilibria, *the presence of the less able individuals lowers the net income of the more able;* in the absence of the second group the first group would have received a wage of θ_1; in the full-screening equilibrium, net income is $\theta_1 - c$, in the no-screening equilibrium it is $\bar{\theta}$. *Conversely, the presence of the more able may increase the income of the less able* (in the no-screening equilibrium, they receive an income of $\bar{\theta}$ rather than an income of θ_2), *but need not* (as in the full-screening equilibrium).

4) *If one of the functions of education is to screen individuals,* as we shall argue later, *social returns* (ignoring distributional effects) *differ from private returns.* The gross social return, in this example, is zero (since the only effects of screening are distributional), the net returns are negative (since there is a cost). But the private rate of return (in the screening equilibrium) to screening, for the more able, is clearly positive:

$$\frac{\theta_1 - \theta_2}{c} - 1$$

Many screening equilibria have the characteristic that some individuals are better off than they would be in the absence of screening, some individuals are worse off, but total net national output is lower. One might be inclined to conclude that such a screening equilibrium is not Pareto optimal, but one must be careful. Assume instead of (1),

$$(1') \qquad \theta_1 - \bar{\theta} > c$$

Then there would not exist a no-screening equilibrium, but the losses from screening to group 2 exceed the gains to group 1. Clearly, if we forbade screening, we could compensate the upper group and divide the costs of screening among the population to make everyone better off. Such an argument misses, however, the essential nature of screening: neither the government nor the private producing sector knows who are the more able without screening; hence, in this example, even though with screening net national output is lower than without it, the screening equilibrium is Pareto optimal in the sense that the redistributions which would be required to make "someone better off without making anyone worse off" than they were in the screening equilibrium, are not feasible in the absence of the screening itself.

---

[2] The multiplicity of equilibria noted here is different from the kind observed by Spence, which arises from an incomplete specification of the equilibrium conditions (or, alternatively, from a different notion of equilibrium than that employed here). (See Rothschild and the author.)

On the other hand, since the screening does lower net national output and increase the inequality of income, ***under any quasi-concave*** (equality preferring) ***social welfare function the screening equilibrium just described is socially undesirable*** (see Anthony Atkinson, Rothschild and the author (1973b)).

These examples illuminate the nature of the private returns to screening: the individual's capturing of his "ability rents" which in the absence of screening he shares with others. It has several special characteristics which are essential for the results: (i) The more able are better in every relevant sense than the less able. Since there is an unambiguous ranking of abilities, we call such screening ***hierarchical.*** (ii) Labor is inelastically supplied and there are no increases in production from sorting individuals. (iii) Individuals have perfect information about their own abilities. (iv) There is no method of on-the-job screening. (v) The screening is perfectly accurate. (vi) The information acquired is "general" information. General information is information about characteristics of an individual which affect his productivity in a wide variety of jobs; specific information concerns characteristics which affect his productivity in a specific firm, for example, his ability to operate a particular machine. (The distinction corresponds to Gary Becker's distinction between general and specific training. These are clearly polar cases; as with training, there is a continuum of degrees of specificity/generality of information.) In subsequent sections I shall show the results are dependent on these characteristics.

In the remainder of this section I consider the private returns to the acquisition of general hierarchical screening information by an individual who is fully informed of his own abilities. I shall argue that in a private economy, as a first approximation, the benefits of such information would accrue to and the costs of information would be borne by the individual as opposed to the firm.

To see this, consider an economy in which individuals did not provide information about themselves. The wage in competitive equilibrium would be equal to the mean marginal product of the workers, and all workers would receive the same wage. Now assume that some firm did research which detected which workers (or groups of workers) were more productive. If it were able to keep that information secret, it would be able to earn, as a return to obtaining that information, the difference between the marginal productivity of these workers and the average of the population as a whole. Thus, it would pay firms to do research to obtain this information, provided, of course, that the costs of obtaining the information were sufficiently low. If the information were to become ***public,*** however, the worker would receive the benefits of the information: other firms would bid for his service, until his wage rose to his marginal productivity.

There are thus two conflicts of interest: the worker wishes to have all such information public, the firm private; and to the extent that some of the return is captured by the worker, the firm will not allocate as much resources to obtaining information about the quality of the individual as the more able individuals would have liked.

We have so far established that the most able individuals have an economic interest in providing information about their capabilities. But the gain of the more productive workers may be at least partially at the expense of the less productive workers. It may be in the interests of the poorer workers for the information about who is the best worker not to be known. I shall now argue that if information were relatively costless, in a competitive economy everyone except the poorest (least capable) individual would have an economic interest in providing such information. For, assume the most able is able to provide information certifying to his abilities. The market would then, in equilibrium, pay the remaining workers their (now lower) mean marginal productivity. It would clearly pay, then, for the most able person of this group to have his ability certified. And the analysis proceeds, until information about the capabilities of all individuals except for the least capable is provided: but if we have sorted out all except for the least capable, we have also sorted out the least capable. This may be called the Walras Law of screening information.

Our basic argument can be summarized as follows: *since individuals are able to capture the returns to general information about their skills themselves, they are willing to spend resources to provide this information*—indeed, this is the only way they can fully capture their "ability rents"; and in a competitive economy, firms that allocate resources to obtaining general screening information about individuals will be unable to appropriate (most of) the returns.

There are some conditions under which even the most able may not be willing to pay for "general screening." (a) If there are self-employment opportunities where they can realize the same returns that they would have realized had they been accurately screened, any "underrated" individual would be self-employed. For most individuals, this is not a relevant possibility. (b) If individuals are perfectly certain of their ability, and if it is possible for their ability to be costlessly observed "on the job" then the individual would offer to absorb all the risk involved in hiring and training costs. There are obviously instances of this sort, individuals who persuade the employer to hire them at low wages until they can "prove themselves." But for many jobs, ascertaining abilities (productivities) on the job may be relatively costly; most individuals are not perfectly certain of their abilities, and the screening is far from perfectly accurate. (c) If individuals are very risk averse and not perfectly certain of their abilities, then they may prefer to be treated simply as average rather than to undertake the chance of being screened and labelled below average. Indeed, in the examples given above, screening increases the variance of the individual's income and reduces the mean (since there is a cost to screening) and so, in such a situation, a "completely uninformed" individual, that is, one who took as the subjective probability distribution of his abilities the distribution of abilities in the population, would never screen. But even if there is a social return to screening, uninformed individuals may not undertake it (see Section ID below).

B. *The Social Benefits from Screening*

The examples of the previous section explicitly assumed that there was no social return to screening; i.e., screening did not increase output, it just redistributed it. Here we discuss the two major categories of social returns.

1. *Tradeoffs.* In the absence of information, individuals receive a wage which differs from their true marginal product. Imperfect information acts just like a wage tax on the more able, a wage subsidy on the less able. Like all taxes, the "information wage tax" is distortionary in its effect on the consumption-leisure decision. If screening costs are small enough, so long as labor is elastically supplied, everyone can be made better off as a result of screening (provided we have the appropriate tax instruments).

(Often, however, the requisite redistributive taxes may not exist; in that case, some of the gains of the more able may be at the expense of the less able.)

Similarly, in choosing a job, an individual must trade off nonpecuniary returns with monetary returns, and if his wage does not correspond to his marginal productivity, he will not make the socially correct decision.

2. *Matching*. Even in the absence of nonpecuniary differences among firms, there is a "matching problem" in the individual's choice of jobs. It is widely recognized that individuals differ in the comparative skills with which they can perform different tasks (jobs) and the ease with which they learn different skills. If the typist has a comparative advantage in plumbing and the plumber a comparative advantage in typing, we can have both more typing and more plumbing if they "switch" jobs.

Educators often talk of the importance of matching an "educational program" to the needs and abilities of our students. The efficiency losses in attempting to train a moron to be an engineer are obvious; other kinds of education mismatching while not as obvious may in the aggregate be quite important.

Even within a given occupation, there are further matching problems. In many economic activities, individuals act together. What is easy to observe is the net output of the group, but this in turn is a complicated function of the different qualities of the individuals of the group. In the previous section, for instance, we considered an assembly line, the speed (output) of which depended simply on the average of the "productivities" of the individuals working on the line. It would perhaps have been more accurate to assume that it is a weighted average, with the individuals who are below average slowing the line down by more than those who are above average speed it up. In that case, total output would be greater if we had two assembly lines, one with slow workers, the other with fast workers, than if the workers were randomly mixed together. Although this example is based on the assumption that there are returns to group homogeneity, the argument that there exist social returns only requires that output depend in part on how individuals of different characteristics are grouped together.

A similar argument can be made with respect to man-machine interactions. Assume that there are different kinds of machines for producing a given level of output. There is a large training cost associated with the operation of each machine; training for one machine does not equip one for operating another. Each machine is optimally designed for an individual of a given ability (value of θ). Clearly there are social returns to knowing the individual's ability (θ). (If there were no training costs, we could quickly observe the output of the machine with any individual, and infer his ability from this.)

C. *Is There Too Little Screening?*

The previous two sections should make it clear that there is no clear correspondence between social and private returns to screening; in the absence of screening individuals are "grouped" together and so may either be subsidized by or be subsidizing other members of the group. Individuals capture the direct increase in their own productivity as a result of screening; but if, as a result of screening, individuals can be "better organized" (for example, by using more homogeneous assembly lines) then there is a kind of externality provided by the availability of information. Moreover, screening eliminates the subsidy which the individual will have been receiving (or extending to others with whom he is grouped). This is a private cost (return) which is not social. As a result of these two factors, there may be too little or too much screening. The following two examples illustrate important situations in which there is too little screening.

1. *Job-Matching Screening: Screening for Comparative Advantage.* Assume a type 1 worker has a productivity of θ_{1s} when assigned to a skilled job but a productivity of θ_{1u} when assigned to an unskilled job. Type 2 workers have a zero productivity on the skilled job. We assume that type 2 workers are actually more productive at the unskilled job than the type 1 workers:

$$\theta_{1s} > \theta_2 > \theta_{1u} \tag{3a}$$

The productivity differentials are such, however, that with no screening, all workers are assigned to unskilled jobs. Let $\bar{\theta}_u$ be the mean wage with no screening in the unskilled jobs,

$$\bar{\theta}_u = h(\theta_1)\theta_{1u} + h(\theta_2)\theta_2$$

Then

$$\bar{\theta}_u > h(\theta_1)\theta_{1s} \tag{3b}$$

If screening costs are such that

$$\max(\theta_{1s} - \bar{\theta}_u, \theta_2 - \bar{\theta}_u) < c < \theta_{1s} - \theta_{1u} \tag{3c}$$

then equilibrium entails no screening; for if an individual of type 1 is screened, his net income is $\theta_{1s} - c$ which is less than his income on the unskilled job, and if an individual of type 2 is screened his net income is $\theta_2 - c < \bar{\theta}$ (again by (3c)). On the other hand, if

$$\max(\theta_2 - \bar{\theta}_u, \theta_{1s} - \theta_2) < c < \theta_{1s} - \bar{\theta}_u < \theta_{1s} - \theta_{1u} \tag{3d}$$

then equilibrium entails a fraction γ of type 1 individuals being screened, where

$$\frac{\theta_2 h(\theta_2) + (1-\gamma)\theta_{1u}h(\theta_1)}{h(\theta_2) + (1-\gamma)h(\theta_1)} = \theta_{1s} - c \tag{4}$$

It is clear that $0 < \gamma < 1$ (at $\gamma = 0$, the right-hand side of (4) exceeds the left-hand side; at $\gamma = 1$, the left-hand side of (4) exceeds the right-hand side). In both cases, net national income maximization entails $\gamma = 1$. Using (3c) and (3d) one can show that by having a subsidy for screening so the cost of screening is lowered to $\theta_{1s} - \theta_2$, financed by a lump sum tax, everyone can be made better off. *If type 1 workers are less productive in unskilled jobs than type 2 workers, there is too little screening.* The reason for this is that in the alternative occupation, the potentially skilled workers are in effect subsidized by the unskilled.

These results do not depend on the lack of complementarity between the two kinds of jobs. For instance, if

$$Q = F(\theta_{1s}\gamma, \theta_2 + (1-\gamma)\theta_{1u})$$

where Q is output and F is a constant return to scale production function, maximization of Q may entail less than full screening but the equilibrium level of screening will still be smaller than the optimal level.

2. *Information Externalities: Returns to Homogeneity.* Assume that the output per worker of the assembly line is of the form

$$\bar{\theta} - \beta\sigma^2 \tag{5}$$

where σ^2 is the variance of abilities on the assembly line. Moreover, assume that there is a fixed, large number of individuals working on the assembly line. Equation (5) embodies the notion that homogeneous work forces work more efficiently. Let σ_θ^2 be the expected variance on the assembly line drawn from an unscreened population, and assume

$$\theta_1 - \bar{\theta} < \theta_1 - \theta_2 < c < \theta_1 - \bar{\theta} + \beta\sigma_\theta^2 \tag{6}$$

Then the (unique) equilibrium involves no screening: with no screening, everyone receives $\bar{\theta} - \beta\sigma_\theta^2$. If a single individual were to buy screening, his income would be (approximately) $\theta_1-\beta\sigma_\theta^2-c$ (since the degree of heterogeneity of the labor force would be unaffected, we assume that the costs of heterogeneity are allocated uniformly over all individuals) which by (6) is less than $\bar{\theta}-\beta\sigma_\theta^2$. On the other hand, with full screening everyone is better off: the lower group receives $\theta_2>\bar{\theta}-\beta\sigma_\theta^2$ (again by (6)) and the upper group receives $\theta_1-c>\bar{\theta}-\beta\sigma^2$. *Although Pareto optimality requires full screening, the market equilibrium entails no screening.* To see that the full-screening situation cannot be sustained by a competitive market (assuming individuals have to pay for their own screening), observe that with full screening the net income of the first group is $\theta_1-c<\theta_2$, the net income of the lower group.

One might have thought that if $c<\beta\sigma_\theta^2$, it would pay firms to screen their workers if they do not screen themselves, since they would then obtain an average output of $\bar{\theta}$ rather than $\bar{\theta}-\beta\sigma_\theta^2$. But if the information about the outcome of screening could not be kept secret (for example, if the two types of assembly lines are different), then type 1 individuals would all be bid away, and so screening would be unprofitable. We assume the firm is aware of this and therefore would do no screening.

D. *Uninformed Individuals*

There is another reason besides the two presented in the previous section why there may not be screening even when it might be possible for everyone to be better off with screening: individuals are uninformed about their abilities and are risk averse. Assume, for instance, that labor is elastically supplied. Then it is possible to show that with the appropriate set of taxes, if the costs of screening are sufficiently small, everyone can be made better off both *ex ante* (expected utility before screening) and *ex post* than in the no-screening equilibrium, but if individuals are sufficiently risk averse, the only equilibrium will entail no screening. The source of "market failure" here is different from those discussed earlier: now the problem is the unobtainability of "ability" insurance, presumably largely because of difficulties with moral hazard.

In such a situation, there is still an incentive for the firm to obtain information about individuals; for if the firm can find individuals whose market wage is below their marginal productivity it can capture the difference between the two, if it can keep the information secret. If, as is often the case, this information cannot easily be kept secret, for example, if individuals of different abilities are assigned to different jobs (kinds of machines), then it would not pay any firm to do screening even if the firm were risk neutral. For other firms would bid away the more productive workers. The firm doing the research would not be able to capture the returns.

There is another problem in competitive economies with uninformed individuals: if two competing firms "discover" that a given individual's marginal product is greater than his wage, then they compete against each other; the individual's wage is bid up until it equals his marginal product, and neither firm is able to capture the returns from doing the research. For a more extended discussion of this point, see the author (1974c).

E. *On-the-Job Screening*

The previous analysis assumed that the screening and production activities were completely separated and there was no on-the-job screening. This is important for two reasons. First, with binding contracts (for the firm not to fire the unproductive, for the productive individuals not to quit), the equilibrium will always be Pareto optimal. For if it were not, any firm, by integrating the screening and production processes could make a pure profit: in effect there is nothing the government could do in these circumstances that an intelligent entrepreneur could not do. In fact, even though there is some on-the-job screening, considerable screening does occur in the educational system, and as long as that is the case, the problems we have detailed above remain. Secondly, on-the-job screening is likely to screen for somewhat different characteristics than, say, educational screening; the return to on-the-job screening is likely to depend on the amount of educational screening and conversely. In the absence of coordination of screening and production, *the equilibrium screening may well be Pareto inefficient,* as the following example illustrates.

Assume individuals are characterized by two characteristics, θ and ϕ, and their productivity is a function of θ and ϕ. (θ may be viewed as a characteristic screened for by the education system, ϕ is a characteristic screened for on the job.) For simplicity, we let $p=p(\theta, \phi)=\theta\phi$. We consider a population with four groups $(\theta_1\phi_1)$, $(\theta_2\phi_1)$, $(\theta_1\phi_2)$, $(\theta_2\phi_2)$, with $\phi_1>\phi_2$, $\theta_1>\theta_2$. Let $h(\theta_i, \theta_j)$ be the proportion of the population with characteristics θ_i and ϕ_j. Define

$$\bar{\theta}(\phi_i) \equiv \frac{\theta_1 h(\theta_1, \phi_i) + \theta_2 h(\theta_2, \phi_i)}{h(\theta_1, \phi_i) + h(\theta_2, \phi_i)}$$

and similarly define $\bar{\phi}(\theta_i)$. Let

$$\text{(7)} \quad \max\left[\phi_i(\theta_1 - \bar{\theta}(\phi_i)), \theta_i(\phi_1 - \bar{\phi}(\theta_i))\right] < c_\theta < c_\phi < \theta_1\phi_1 - \sum_i \sum_j \theta_i\phi_j h(\theta_i, \phi_j)$$

where c_θ and c_ϕ represent the costs of screening for θ and ϕ, respectively. Costs are assumed to be such that it always pays to screen for one and only one characteristic. It immediately follows from (7) that there may be two equilibria, one in which θ is to be used as the "screen," the other in which ϕ is used. When ϕ is being used as a screening device, it does not pay to use θ, and when θ is used, it does not pay to use ϕ. Clearly national income is higher if the former is used rather than the latter. Indeed, it is even possible to construct examples[8] in which everyone is worse off in the former equilibrium rather than the latter! *An attempt to eliminate educational screening may just shift the focus of screening, and make everyone worse off.*

F. *Accurate Screening and Fines*

Another important implication of the possibility of on-the-job screening at any finite cost is that if it is perfectly accurate and individuals are perfectly informed, *the market equilibrium will be characterized by full screening without spending any resources on screening.* The individual agrees to pay the firm a large fine if it turns out he has overstated his ability. The firm announces it will undertake screening of individuals on an assembly line if the output of that assembly line differs from what it should be, given the ability levels which the individuals have declared. Clearly, for a sufficiently high fine, only individuals of ability level θ_1 will declare themselves to be of ability θ_1, and hence no screening need actually be undertaken.

This type of screening often occurs, although in a slightly modified form. Individuals accept low wages while they prove themselves; the low wages today are compensated for by high wages later if they do prove themselves. If they do not, the difference between the low wages and what they could have obtained elsewhere acts as a fine (see Section IIB below and Salop and Salop). Lack of knowledge about one's own abilities and imperfectly accurate screening, combined with risk aversion, places a limit on the efficacy of this kind of screening; if screening is to occur, there will have to be some expenditures for examination. (See the author (1974c).)

---

[8] Let $\theta_1=\phi_1=2$, $\theta_2=\phi_2=1$; $h(\theta_1, \phi_1)=h(\theta_2, \phi_1)=1/3$, $h(\theta_1, \phi_2)=h(\theta_2, \phi_2)=1/6$. Let Y_{ij} be gross income of someone with characteristics (θ_i, ϕ_j). Then with screening for θ: $Y_{11}=Y_{12}=10/3$, $Y_{21}=Y_{22}=5/3$; with screening for ϕ: $Y_{11}=Y_{21}=3$, $Y_{12}=Y_{22}=3/2$. If $C_\theta-1/3\leq C_\phi$, $C_\theta\leq 11/6$, and $C_\phi\geq 4/3$, all individuals are better off under θ screening than under ϕ screening. For these to be equilibria, we require in addition, $1\leq C_\theta\leq 5/3$, and $C_\phi<3/2$. With the further restriction that $C_\theta+C_\phi>9/4$, it can be shown that there are no other equilibria.

G. *Nonexistence of Equilibrium*

We have exhibited examples of too much screening, too little screening, the wrong kind of screening, and multiple equilibria. But another striking aspect of screening models is that ***there may be no competitive equilibrium*** where individuals take the action of others as well as the wages paid to an individual of any label as given.

The simplest example involves a slight modification of the one given in Section IE. For simplicity we present only a numerical version: Let $p(\theta_1, \phi_1)=4$, $p(\theta_2, \phi_2)=2$, $p(\theta_2, \phi_1)=p(\theta_1, \phi_2)=0$, $h(\theta_i, \phi_j)=\frac{1}{4}$, all i, j; let $c_\phi=c_\theta=1.5$. Clearly, there exists no no-screening equilibrium $(4-1.5>1.5)$.

Let us consider alternative possible screenings. Assume $(\theta_1\phi_1)$ screens for θ only. It then pays a fraction (approximately .7) of $(\theta_1\phi_2)$ to screen for θ_1. But this cannot be an equilibrium, for the average wage (after paying for screening) of those screened for θ_1 is then .9; clearly, it pays $(\theta_1\phi_1)$ to screen for ϕ as well. (His net income would then be 1.) But if $(\theta_1\phi_1)$ screens for both θ and ϕ, it does not pay $(\theta_1\phi_2)$ to screen for θ. But if $(\theta_1\phi_2)$ does not screen, it does not pay $(\theta_1\phi_1)$ to screen for both θ and ϕ. Other possibilities (for example, $(\theta_1\phi_1)$ screening for ϕ, partial screening, etc.) may be checked, to see that there in fact exists no equilibrium. (This is similar to the result of Rothschild and the author (1973a).)

II. Screening and Education

Section I established some general characteristics of screening equilibria. We now focus in more detail on screening in educational institutions. Educational institutions are not the only institutions which do screening in our economy. Employment agencies and the College Entrance Examination Board both screen; there is considerable on-the-job screening; how an individual dresses, his accent, his socioeconomic background, his race or ethnic group may all provide bases for screening. The fact that there are other bases for screening does not detract from the importance of educational screening; indeed the screening done by educational institutions provides the primary determinant of one's initial job opportunities and hence of what screening can occur subseqeuntly. In this section we enquire into why educational institutions are important for screening (Section IIA), the mechanisms used for screening (Section IIB), and the implications this has for the structure of the educational system (Sections IIC–E).

A. *Why Educational Institutions?*

Educational institutions provide information about individuals' abilities for a number of reasons: (a) The efficient allocation of scarce educational resources requires the identification of different individuals' abilities, i.e., some individuals would gain little from a Ph.D. program in economics, but would clearly benefit greatly from a course in automobile mechanics, and conversely for other individuals. (b) Most educators would argue that even within a given educational level there are returns from recognizing that some individuals learn certain skills faster than others. (c) Part of the social marginal product of educational institutions is finding each individual's comparative advantage (as educators are wont to say, "helping the individual find out about himself") and information about absolute advantages is almost an inevitable byproduct of obtaining information about comparative advantages. (d) In the interchange between teacher and student which is common to many (but not all) educational processes, the teacher obtains a great deal of information about his student. The fact that there are a large number of teachers making those "observations," makes the information more valuable than the judgment of a single individual (for example, an employer).

In short, it is hard to imagine an educational system which did not obtain some information about individuals. Not all educational *processes* involve screening; that is, large lectures may impart a great deal of information, but the teacher need never ascertain how much of the information the student has absorbed. Some students have even argued that screening diverts them from "real" education to the acquisition of the particular skills and pieces of information which will be tested. Our analysis is predicated on the fact that for the reasons mentioned above, all educational *systems* do some screening.

B. *The Provision of Screening Information: The Screening Mechanisms*

As discussions of grading systems make clear, there is, however, an important difference between obtaining information and making it public. There are several mechanisms by which such information about the individual's capabilities become public:

1) If the education system does any sorting for its own purposes (as it must), the groups into which an individual has been sorted will convey some information to the firm about the individual.

2) Another mechanism is performance tests: individuals have been confronted with roughly similar learning experiences (say geometry). Some individuals "learn" geometry better than others: this fact may be ascertained by a "grade" from the teacher, or by "standardized" objective examination. Failure to pass a course in college, or failure to pass a grade in elementary and secondary schools, conveys a great deal of information, which adversely affects the wages received by those individuals. As long as the school system does any grading, if only on a pass-fail basis, it is providing some information; and even when it does not do the grading itself, others can do the grading for it (Graduate Record Examination, etc.).

3) A great deal of information is provided, however, by *self-selection:*[4] a self-selection mechanism works as follows. Consider any characteristic of an individual about which the individual has more information than the firm. (We do not require that the individual have perfect information, only that on average he be better informed than the firm.) Some individuals have "more" of the given characteristic than others, for example, more brains, more mechanical ability, a higher turnover rate. We construct two (or more) reward-penalty structures such that on average individuals with more of the given characteristics will do better under one reward-penalty structure than under the other, and conversely. If individuals are asked to choose among these reward-penalty structures, and if they are rational, they will sort *themselves* out into those who have more of the characteristic and those who have less. (The better the information of the individuals and the greater the differential rewards, the better the sorting will be.)

Assume that wages are a function of the number of grades completed, and the length of time to complete a grade is a function of the individual's ability. Then if the two functions have the appropriate shape, individuals with lesser ability will quit at a lower grade level than persons with a higher ability. Grade completed is a complete surrogate for ability (see Spence).

---

[4] This is related to Akerlof's theory of lemons (1970). Akerlof argues that the used car market is a self-selection mechanism in which the worst cars become traded. Self-selection mechanisms provide what Spence has called "signals" to the market.

Alternatively, assume we have a hierarchy of schools, from those for the most able to those for the least able. Assume that the schools only use a pass-fail system. Assume that the schools for the more able are more expensive. If individuals had perfect information about their capabilities (and ignoring motivation, emotional, and other problems) in fact no one need ever fail. Students would apply to the school of the appropriate ability.

It should be noted that all these self-selection devices are based on *performance tests;* that is, although the employer is using information from self-selection, self-selection only works because of the performance tests. If there were no *possibility* of failures, everyone would attempt to go to the best school (and then screening would have to be done by admissions committees) and everyone would pass on from grade to grade at the same rate.[5]

C. *The Structuring of Educational Systems*

Although we have argued that an educational system inevitably provides some information about the capabilities of individuals, there are a number of characteristics of the school system which determine how much and what kind of information is provided either by performance test or by self-selection. The school system can decide on the fineness or coarseness of screening. The structure of payments for education and the differences in "levels of education" provided by different schools are also important determinants of the effectiveness of self-screening.

Earlier, we noted that the reason that the school system is the major screening institution in our society is that this information is a natural by-product of its principal activity of providing knowledge (skills) and guiding individuals into the right occupations. In most of the ensuing analysis, we shall employ a stronger hypothesis: the more educational institutions perform their principal functions, the more screening that is produced as a by-product.[6] The more accurately it is able to place individuals into the right "slots," i.e., ascertain their comparative abilities, the more accurately it must ascertain the individuals' absolute abilities. The more knowledge it attempts to impart, the more it is able to "separate the men from the boys." At the extreme, if it tried to teach nothing, there would be no basis for performance testing, and there would similarly be no basis on which the self-screening mechanisms could be based.

There is thus the possibility that in imparting more skills to the abler students, we will simultaneously increase the inequality of income. This has made the organization of the educational system, and the method by which the levels of screening and skill acquisition are determined, an intensely political question.

Many of the social issues involving education arise because of differences in the wealth of parents. It is important, however, to observe that this parental distributional question can at least partly be separated from the questions of educational organization on which we are focusing. Thus the government could provide its support for education in the form of vouchers, allowing individuals to use these in private schools. Even if there were no inequality in parental ability to pay for education, there would be, as we have argued above, important distributional consequences to alternative methods of organizing the educational system. To isolate our attention on these, we shall assume in the subsequent discussion that an individual's attitude towards education is determined completely by the own private monetary returns.[7]

---

[5] This is, of course, not true of other self-selection mechanisms, e.g., those discussed by Salop and Salop. The absence of performance tests plays a crucial role in the economics of self-selection devices discussed by Akerlof (1973), and Rothschild and the author (1973b).

[6] That is, for most of the analysis we shall assume that they are joint products, and that the mix between screening and "skill formation" is technologically determined. We could generalize the model to allow for the determination of this mix. In this paper we will not enquire in detail how the skill acquisition and screening take place (e.g. the nature of the grading system). We shall employ a general formulation which is consistent with a number of alternative microstructures.

D. *The Comprehensive School Systems with Majority Voting and Fairly Accurate Screening*

In this section, we shall show that *with majority voting a comprehensive school system will under reasonable assumptions allocate too many or too few resources to education (screening), relative to the amount which would maximize net national output depending on whether individuals are informed or uninformed about their abilities.*

The model is a slight extension of that presented in Section 1. Individuals are described by a single characteristic θ; the distribution of θ over the population is given by $h(\theta)$. We let λ denote the "intensity" of education.[8] More intensive education (a) costs more, (b) screens better, and (c) increases the productivity of the group educated, either because of skill acquisition or better matching of individuals and jobs.

The Productivity Effect. Let $p(\theta, \lambda)$ be the productivity of an individual of ability θ who has received an education of intensity λ. For simplicity, we shall let p take on the special form (upon appropriate choice of units)[9]

$$(8) \qquad p(\theta, \lambda) = m(\lambda)\theta, \quad m' \geq 0, \quad m'' \leq 0$$

Screening. The educational system places labels on individuals; it gives a point estimate of the individual's ability. Let $e(\theta, \hat{\theta}, \lambda)$ be the probability that an individual of type θ be labelled $\hat{\theta}$, in an educational system of intensity λ. As λ increases, the probability of error decreases, i.e.,

$$(9) \qquad \left.\frac{\partial e(\theta, \hat{\theta}, \lambda)}{\partial \lambda}\right|_{\hat{\theta}=\theta} \geq 0$$

Costs of Education. Finally, we assume that the cost of education per pupil $c(\lambda)$ is an increasing function of λ and that the marginal cost also increases with λ.

$$(10) \qquad c' > 0 \quad \text{and} \quad c'' > 0$$

In a comprehensive educational system all schools have the same value of λ. The model includes as special cases the traditional model of pure skill acquisition ($\partial e/\partial\lambda = 0$) and the pure screening model ($m' = 0$).

Wage Determination. Workers whose ability is estimated to be $\hat{\theta}$ receive a wage equal to their mean marginal product

$$(11) \qquad w(\hat{\theta}) = m(\lambda) \int \theta e(\theta, \hat{\theta}, \lambda) h(\theta) d\theta \div \int e(\theta, \hat{\theta}, \lambda) h(\theta) d\theta$$

---

[7] This would be the case for instance even without government redistribution if (a) there were a perfect capital market, (b) education were not a consumption good, and (c) there were no tax distortions in the allocation of capital between human and physical capital.

[8] Throughout the discussion we make the extreme assumption that all information about individuals' abilities is obtained through the educational system, and hence the individuals' wages are determined by the label imposed by the schools. Obviously, there is some information obtained on the job. The qualitative results of our analysis will, however, be unaffected so long as (a) firms cannot obtain information on the job instantaneously, and/or (b) there are any fixed costs of hiring and training. Intensity can be thought of as either "length" (number of years of schooling) or "quality" within a program of fixed length.

[9] It should be noted that the model may be considerably generalized without affecting its qualitative properties. In particular, the restriction embodied in equation (8) may be dropped, and an additional kind of education which increases skills without screening may be introduced. See the author (1972b).

The expected wage which a person whose true ability is θ will receive is then given by

$$W(\theta) = \int w(\hat{\theta})e(\theta, \hat{\theta}, \lambda)d\hat{\theta} \tag{12}$$

We shall consider the special case of a fairly accurate grading system in which $e(\theta, \hat{\theta}, \lambda)$ takes on the form

$$e(\theta, \hat{\theta}, \lambda) = f(\theta - \hat{\theta}, \lambda) = f(\epsilon, \lambda) \tag{13}$$

where $\epsilon = \theta - \hat{\theta}$ is the error. We thus assume that the distribution of error is independent of the value of θ. Moreover, we assume $E\epsilon = 0$ and $E\epsilon^2 = g(\lambda)$, $g'(\lambda) \leq 0$. Thus from (11)

$$w(\hat{\theta}) = m(\lambda)\hat{\theta} + \frac{m(\lambda)\int \epsilon f(\epsilon, \lambda)h(\hat{\theta} + \epsilon)d\epsilon}{\int f(\epsilon, \lambda)h(\hat{\theta} + \epsilon)d\epsilon} \tag{14}$$

$$\simeq m(\lambda)\left[\frac{h'(\hat{\theta})g}{h(\hat{\theta})} + \hat{\theta}\right]$$

(15)

$$W(\theta) \simeq m(\lambda)\int\left[\hat{\theta} + \frac{h'(\hat{\theta})}{h(\hat{\theta})}g\right]f(\theta - \hat{\theta}, \lambda)d\hat{\theta}$$

$$= \int\left[m(\lambda)(\theta - \epsilon) + m(\lambda)\frac{h'(\theta - \epsilon)}{h(\theta - \epsilon)}g\right]f(\epsilon, \lambda)d\epsilon$$

$$\simeq m(\lambda\left[\theta + \frac{h'(\theta)}{h}g\right] \gtrless m(\lambda)\theta \text{ as } h' \gtrless 0$$

Thus in an unimodal distribution, individuals below the mode get more than they would under perfect screening, individuals above the mode get less than they would. The reason for this is that individuals are being averaged with some individuals who are better than they are, but have been underrated, and some who are worse, but who are overrated; if there are more who are worse (within a given range of error) than who are better, the individual will receive less than his true marginal productivity (on average).

Output Maximizing Educational Intensity. If we wish to maximize national output less educational expenditures, i.e.,

$$\max\left\{m(\lambda)\int \theta h(\theta)d\theta - c(\lambda)\right\} \tag{16}$$

we set

$$c'(\lambda) = \bar{\theta}m'(\lambda) \tag{17}$$

where $\bar{\theta}$ is the mean level of ability in the economy. The solution to (17) we shall call the "optimal level of education," bearing in mind that we are using the term in a very restricted sense.

Majority Voting. We now come to the choice of an educational intensity (and the associated degree of screening) in a majority voting political system. We assume the educational system is paid for by proportional wage taxes. Then if τ is the tax rate,

$$\tau m(\lambda)\bar{\theta} = c(\lambda) \tag{18}$$

and the net expected wage of someone at ability θ (using (15) and (18)) is

$$W(\theta)(1 - \tau) \tag{19}$$

$$\simeq m(\lambda)\left[\theta + \frac{h'}{h}g\right]\left(1 - \frac{c(\lambda)}{m(\lambda)\bar{\theta}}\right)$$

$$= \left(\theta + \frac{h'}{h}g\right)\left(m(\lambda) - \frac{c(\lambda)}{\bar{\theta}}\right)$$

Taking the derivative of (19), we can see how varying educational intensity affects different groups

$$\frac{dW(1 - \tau)}{d\lambda} = \frac{W}{m}\left(m' - \frac{c'}{\bar{\theta}}\right) + (1 - \tau)m\frac{h'}{h}g'$$

This depends on both θ and λ. Consider the optimal level of education. Note then that the first term drops out, and we are left with only the second term: individuals above the mode will want more than the optimal level of education, individuals

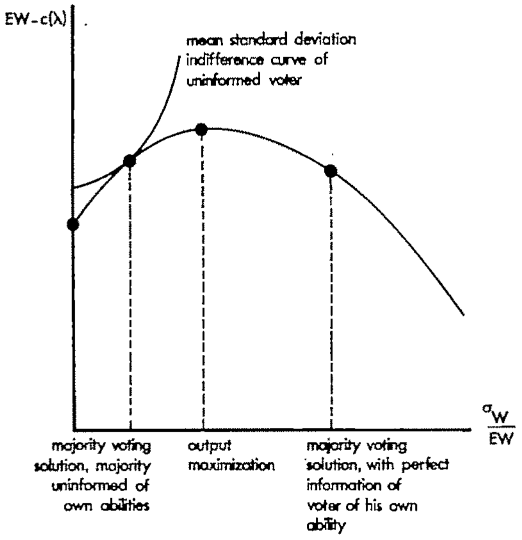
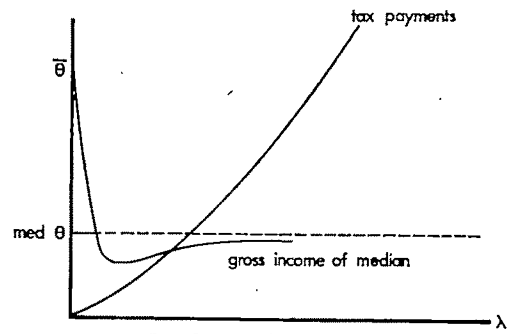
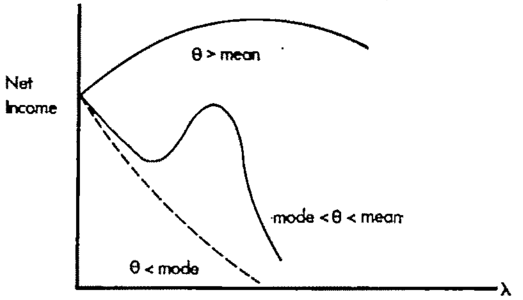


FIGURE 1



Pure Screening (no productivity effects)



Screening (with productivity effects)

FIGURE 2

below the mode will want less. It is possible to show that if c'', $g''\geq 0$, and $m''<0$, preferences will be single peaked. Thus, the majority decision will be determined on the basis of the median value of h'/h. It is clear that if the mode lies below the median as it does for the income distribution, there will be an excess of investment in education over the optimum amount.

Indeed, it is easy to establish that not only is output lower, but the coefficient of variation in after tax expected wage income is greater, as illustrated in Figure 1.

It is worth noting at this point a major difference between fairly accurate screening systems and those which, for low values of λ, are very inaccurate. Take as an extreme case a system in which with "no information" and no education everyone receives the average value of the marginal product, as discussed earlier. Assume education only screens and that the distribution of abilities is lognormal. With no screening, the median receives the average, with perfect screening, he receives the median. As screening increases, his gross income initially declines. The cost of education increases with screening. Thus, there are two "peaks" to his net income, and accordingly there may not exist a majority voting equilibrium.[10] (See Figure 2.)

The above analysis assumed that everyone knew perfectly his own ability. The other polar case is where at least a majority of individuals are completely uninformed as to their abilities, i.e., their subjective probability distribution of their abilities is identical to the frequency distribution of abilities in the population. It is clear then that the median voter will vote for a level of education which is below that which maximizes net national output.

E. *Noncomprehensive School Systems*[11]

Although there is an institutional and analytical simplicity to a comprehensive school system, it is easy to establish that in general net national output is not as high as in a system in which different individuals receive a different education. Indeed, if by greater ability we mean in part the ability to learn more easily, then it is more efficient (if our objective is maximizing net national output) to spend more resources on the more able.[12] This will be a characteristic of most noncomprehensive school systems. The allocation, however, will differ between a governmentally organized system attempting to maximize net national output, a private educational system, and a mixed public-private system. A full analysis would take us beyond the scope of this paper, but what we wish to do here is to characterize the major reasons that the equilibrium in pure private as well as mixed public-private systems does not maximize net national income.

---

[10] I am indebted to John Chant for discussions on these points.

[11] For a more extensive discussion of the issues discussed here as well as the development of a formal model, the reader is referred to the author (1972b, 1974).

For simplicity, it is best to return to the special case of Section IA, where there are only two ability groups in the population. The school system will consist of two schools, one run for the more able, one for the less able. In the mixed public-private school system, the school for the more able is private, for the less able, public. Private schools charge a tuition equal to per pupil expenditure; public schools raise revenue by general proportional taxation. We assume that the less able are in the majority. Each school system will have some of both kinds of individuals, the upper school will contain some individuals of lower ability who are attempting the "gamble" of being able to pass through the system and hence be grouped with the more able, and those of lower ability who overestimate their ability. Conversely for the lower school.

We shall now argue that there is some presumption for excessive expenditure even in a private school system. Consider the three effects of an increase in educational expenditure in the upper school. First, there is the direct productivity effect. Since the upper school focuses its attention on those who will "succeed," it spends more on this account than a government-run school which is also concerned with those who do not succeed in the upper school. Secondly, there is the direct screening effect, which, as we argued above, is simply redistributive in character, and again leads to "too much" spending on education. Thirdly, there is the "self-selection effect." By increasing educational expenditure and the quality of screening, the upper school discourages those of lower ability from attempting to go to the upper school. There is some social return to this, since the amount of education which is optimal for the less able is less than that which is optimal for the more able. The private return, however, is derived not from the increased "efficiency" of the educational system, but from the ability of the more able to capture more of their "ability rents." The private return to self-selection may be more or less than the social return. Thus, only if the social return to self-selection exceeds the private return by just the right amount to compensate for the excess of the private productivity and direct screening returns over the social returns will the level of expenditure be at the output maximizing level; normally we would expect there to be too much expenditure in the upper school.

A similar analysis applies to the lower school. It is obviously not in interests of those of lower ability to have extensive screening. Although the social return to self-selection is positive, the private return to those of lower ability is negative. By increasing the level of educational expenditures they are, however, able to attract those of higher ability who are less sure of their abilities and more risk averse.

---

[12] Although the precise quantitative relationship clearly depends on the specific technological assumption embedded in equation (8) so long as some are able to learn more quickly and easily than others, the result remains valid.

This again leads to some presumption of excess spending even in the lower school. When the lower school is publicly financed, there is a further incentive for excess spending, since now the costs for the lower school are borne by the population as a whole.

F. *Concluding Comments*

In recent years economists have shown an increasing awareness of "market failures" and have increasingly called upon government intervention to correct these failures. But to turn over an allocation process to the public sector is to make it subject to "political laws" which may be no less forceful—and even less efficient—than the "economic laws" which previously governed the allocation process. The fact that these political laws are less well understood, perhaps more amorphous, than the corresponding economic laws is not an excuse for relying on the mythical "benevolent despot" who plays the central role in most economists' models of the public sector.

The educational sector provides an important point of comparison between the two allocation processes. If, as we have suggested, education provides information as well as skills, then it is providing a "commodity" for which it is well known that the market "fails"; we have shown how social returns differ from private returns and have examined in detail the market allocation of resources to education as well as the structure of the educational system which would emerge from a simplified political process in a highly idealized setting. Some important results emerge. Screening has productivity returns, but tends to increase inequality. There will thus be a tradeoff between efficiency and distributional considerations; but beyond a certain point, further increases in educational expenditure may both increase inequality and decrease net national income. We noted a tendency for all the school systems examined—public, private, and mixed—to operate at these levels even when all citizens are simply concerned with their own income maximization. One of the reasons for this—found in all of the systems—is that some of the returns to higher levels of education (those returns derived from the increased accuracy of labelling individuals' abilities), are private but not social returns; we argued that if abilities are distributed skewly to the right, for the median voter these private returns were positive. A further reason, in publicly supported systems, is that the median voter pays for less than his proportionate share in marginal costs. As a result, the tendency for excessive spending on education may be greater in the publicly financed schools.

On the other hand, it should be emphasized, that whether there is "too much" or "too little" screening in a competitive economy depends on a number of assumptions concerning the screening technology, how well-informed individuals are concerning their own abilities, the nature of the production process, and whether screening is primarily hierarchical or "job-matching."

Finally, we note that attempts to curtail educational screening may simply shift the focus of screening (for example, to on-the-job screening), with the possibility of a lowering of net national output without any commensurate gain in equality.

REFERENCES

G. Akerlof, "The Market for 'Lemons': Qualitative Uncertainty and the Market Mechanism," *Quart. J. Econ.*, Aug. 1970, *89*, 488–500.

———, "A Theory of Information and Labor

Markets," paper presented at the *NSF-NBER* Conference on the Economics of Information, Princeton 1973.

K. Arrow, "Models of Job Discrimination" and "Some Mathematic Models of Race in the Labor Markets," in M. H. Pascal, ed., *Racial Discrimination in Economic Life*, Lexington 1972.

———, "Higher Education as a Filter," *J. Publ. Econ.*, July 1973, *2*, 193–216.

A. B. Atkinson, "On the Measurement of Inequality," *J. Econ. Theory*, Sept. 1970, *2*, 244–63.

G. Becker, *Human Capital*, New York 1964.

G. Fields, "Towards a Model of Education and Labor Markets in Labour Surplus Economies," mimeo., Growth Center, Yale Univ. 1972.

J. Hirschleifer, "The Private and Social Value of Information and the Reward to Inventive Activity," *Amer. Econ. Rev.*, Sept. 1971, *61*, 562–74.

R. Hull and L. Peters, *The Peter Principle*, Suffolk, Great Britain 1969.

E. Phelps, "The Statistical Theory of Racism and Sexism," *Amer. Econ. Rev.*, Sept. 1972, *62*, 659–61.

M. Rothschild and J. E. Stiglitz, (1973a) "Equilibrium in Competitive Insurance Markets: The Economics of Markets with Imperfect Information," paper presented at the *NSF-NBER* Conference on the Economics of Information, Princeton 1973.

——— and ———, (1973b) "Some Further Results on the Measurement of Inequality," *J. Econ. Theory*, Apr. 1973, *6*, 188–204.

S. Salop, "The Noisy Monopolist: Imperfect Information, Price Dispersion and Price Discrimination," unpublished paper, Fed. Res. Board, Washington 1973.

J. K. Salop and S. C. Salop, "Self Selection and Turnover in the Labor Market," paper presented at Econometric Society Meeting, Toronto 1972.

M. Spence, "Job Market Signalling," *Quart. J. Econ.*, Aug. 1973, *87*, 355–79.

J. E. Stiglitz, "Perfect and Imperfect Capital Markets," paper presented at Econometric Society Meeting, New Orleans 1971.

———, (1972a) "Equilibrium Wage Distributions," mimeo., 1972.

———, (1972b) "Education as a Screening Device and the Distribution of Income," mimeo., Yale Univ. 1972.

———, "Conceptual Approaches to the Economics of Discrimination," *Amer. Econ. Rev. Proc.*, May 1973, *63*, 287–95.

———, (1974a) "Demand for Education in Public and Private School Systems," *J. Publ. Econ.*, forthcoming.

———, (1974b) "Theories of Discrimination and Economic Policy," in G. von Furstenberg, ed., *Studies in the Economics of Discrimination*, Lexington 1974.

———, (1974c) "Information and Economic Analysis," *AUTE Conference Proceedings 1974*, Cambridge forthcoming.

L. Thurow, "Education and Economic Equality," *Publ. Interest*, Summer 1972, *20*, 61–81.

M. Young, *The Rise of the Meritocracy*, Baltimore 1958.

Substitution Effects and Biases in Nontrue Price Indices

By P. J. LLOYD*

A number of indices of prices such as the Laspeyres and Paasche are commonly used to compare price movements between two or more situations, but they are not true price indices. When the commodities purchased are final consumer goods, the true price index, or the true "cost of living" index as it is often called, is defined by Paul Samuelson and S. Swamy as ". . . the ratio of the (minimum) costs of a given level of living in two price situations" (p. 567). One may define a true cost of production index for a single commodity or a fixed bundle of commodities, or for national production possibilities (see the survey by Samuelson and Swamy), in an analogous way. Unfortunately, the quantities of commodities that would be purchased in at least one situation are not directly observable.

Two approaches to the measurement of true indices have been developed. The first is to establish upper and lower limits, preferably in terms of price indices using observable data, within which the true measure must lie; Ragnar Frisch and Samuelson (1947, pp. 156–63) survey several known limits. The second is to make direct estimates of true price indices. These have been made, so far as I am aware, only for the rather special case of consumer price movements on the maintained hypothesis that the consumer's utility function[1] is Stone-Geary (see Arthur Goldberger and Theodore Gamaletsos, and Tran Van Hoa). I shall extend this second approach.

Section I presents the true price index in terms of the minimum expenditure function. In addition to yielding, by definition, direct estimates of the true price index, this function also provides a relatively simple way of exploring its properties.

Two special properties of these indices are considered under the assumptions that the utility functions are homothetic or homogeneously separable. As an illustration, true cost of living indices are derived for a number of particular utility (production) functions. In Section II one of these functions, the two-level *CES* function of Kazuo Sato, is used to find the magnitude of the bias in the estimates of true price indices obtained by employing Laspeyres or Paasche indices. Contrary to the indications of the restrictive Stone-Geary function, the bias can be substantial for realistic parameter values and price movements. In Section III, I indicate how the methods used may be extended to incorporate leisure, multiperiod consumption and saving, and durable goods in the model. The paper will not cover other problems of index construction, such as those posed by changes in tastes (see

---

* Australian National University. I am indebted to Ted Sieper, Sundrum, Stephen Turnovsky, and a referee for comments on an earlier draft. Section I has been revised and condensed, taking advantage of the survey by Paul Samuelson and S. Swamy published earlier in this *Review*.

[1] A third approach is to derive an approximation to the price index by assuming some linear, quadratic, or higher-order approximation to the iso-utility surface (see Frisch, pp. 27–31, for several earlier examples) or to the ordinary commodity demand functions (see Lawrence Klein and Herman Rubin) or to total expenditures (see Henri Theil). See also Samuelson and Swamy.

Samuelson and Swamy), or qualities of commodities (see Franklin Fisher and Karl Shell), the presence of public goods, and the need to aggregate across individual buyers.

I

We are concerned with obtaining for a single consumer the true price index from a utility function which is assumed to represent in an order-preserving way his binary preference orderings and which does not vary over situations. There are n commodities. All results carry over to the analogous problems in production theory.

The consumer acts to maximize his utility function

$$U = U(q_1, \dots, q_n) \tag{1}$$

subject to the budget constraint

$$\sum_i p_i q_i = m, \qquad m = \text{constant}$$

The true price index measuring the change in the level of prices between two situations, situation 0 and situation 1, is defined as

$$T = e^1/e^0 \tag{2}$$

where

$$e = \min \left\{ \sum_i p_i q_i \,\middle|\, U(q_1, \dots, q_n) = \mu \right\}$$

for a given set of prices. It is immediately apparent that the true price index is simply the ratio of two values of the minimum expenditure or support function[2] (see Ronald Shephard (1953, ch. 3) or Lionel McKenzie). The function e is linearly homogeneous in its price arguments. From the definition of e

$$T = \overset{1}{e}(\overset{1}{p}_1, \dots, \overset{1}{p}_n; U)/\overset{0}{e}(\overset{0}{p}_1, \dots, \overset{0}{p}_n; U) \tag{3}$$

That is, the true price index is in general a function of prices in situations 0 and 1 and of the chosen level of utility. It is customary but not necessary to take the level of utility in situation 0 as the level of utility to which the two minimum expenditures relate. In this case

$$T = \sum_i \overset{1}{p}_i \overset{*}{q}_i \Big/ \sum_i \overset{0}{p}_i \overset{0}{q}_i \tag{4}$$

where now the denominator is the actual expenditure in situation 0 under the assumption of utility maximization; that is, $e^0 = m^0$. The numerator is the minimum total expenditure required to attain the level of utility in situation 0, given the set of prices in situation 1. We shall use this case throughout as an illustration. In this case $(q_1^*, \dots, q_n^*)$ for situation 1 cannot be observed from market behavior.

The index may be written as the weighted arithmetic mean of ratios of minimum expenditures for each commodity, or for groups of commodities. From the definition

$$T = \sum_{j=1}^{m} \overset{1}{e}_j \Big/ \sum_{j=1}^{m} \overset{0}{e}_j \qquad m \leq n \tag{5}$$

$$= \sum_{j=1}^{m} (\overset{1}{e}_j/\overset{0}{e}_j) w_j \qquad \sum_j w_j = 1 \tag{6}$$

where $e_j = p_j q_j$ is the cost-minimizing expenditure on commodity or commodities j and $w_j = e_j^0 / \sum_j e_j^0 > 0$ is the share of commodity j in the total expenditure of situation 0. In general the quantity of each commodity and therefore each e_j will depend on the prices of all commodities. These ratios of expenditures are not the price relatives (p_j^1/p_j^0), except in the case of fixed consumption proportions (see equation (26) below). T may also be writ-

---

[2] There is an alternative method of derivation of T from the indirect utility function $U = \psi(p_1, \dots, p_n; e)$ which has been used by Goldberger. For a chosen U,

$$U = \psi(\overset{0}{p}_1, \dots, \overset{0}{p}_n; \overset{0}{e}) = \psi(\overset{1}{p}_1, \dots, \overset{1}{p}_n; \overset{1}{e})$$

we may solve for e^0 and e^1 which are as before, the minimum expenditure to attain U, given prices in the two situations. The duality of the indirect and direct utility functions has also been used for homothetic functions by Samuelson and Swamy.

ten as a general mean of the set of all price relatives $(p_1^1/p_1^0, \ldots, p_n^1/p_n^0)$.

When further restrictions are imposed on U or ψ, the form of the true price index itself may be further restricted. One basic question is: Under what restrictions can we obtain an index in which m^0 does not appear as an argument, that is, in which the true price change is independent of the initial level of utility and expenditure?

THEOREM 1: (*The Bergson-Shephard Theorem*);[3] *The true price index may be written in the form*

$$(7)\quad T = P(\overset{1}{p}_1, \ldots, \overset{1}{p}_n)/P(\overset{0}{p}_1, \ldots, \overset{0}{p}_n)$$

where $P(p_1, \ldots, p_n)$ *is homogeneous of degree one, if and only if the utility function,* U, *is homothetic.* (For proof of this well-known theorem see Shephard (1970) or Samuelson and Swamy.) Homotheticity gives utility functions the special geometric property that all level surfaces can be obtained by radial expansion from the origin of the unit level surface. Given a set of initial prices and hence a cost-minimizing set of commodity proportions, a new set of commodity prices must result in a certain proportionate change in expenditure to maintain the initial level of utility, and therefore, a certain value of T, independent of the initial level of expenditure. This family of utility functions is free of the objection that the true price index is arbitrary because it applies only to a particular level of utility (see R. G. D. Allen, pp. 289–90).

Another interesting question is the conditions under which the true price index may be written as some function of sub-indices for the prices of groups of commodities, and therefore, the movement of prices between two or more periods may be decomposed into the movements due to the prices of the separate groups of commodities. This is not, in general, possible but a condition that is sufficient (but not necessary)[4] may be derived readily from a well-known aggregation theorem.

THEOREM 2: *The true price index may be written in the form*

$$(8)\quad T = e(\overset{1}{\bar{p}}_1, \ldots, \overset{1}{\bar{p}}_m; \overset{0}{U}) / e(\overset{0}{\bar{p}}_1, \ldots, \overset{0}{\bar{p}}_m; \overset{0}{U})$$

where $\bar{p}_j = \bar{p}_j(p_1^j, \ldots, p_{n_j}^j)$, $j = 1, \ldots, m < n$; $\sum_{j=1}^m n_j = n$ *if the direct (or indirect)*[5] *utility function is homogeneously separable.* Note that $\bar{p}_j$ is the price for the jth group of commodities $(q_1^j, \ldots, q_{n_j}^j)$. A utility function is said to be homogeneously separable with respect to a partition of the n primary commodities into m mutually exclusive and exhaustive subsets if and only if:

$$(9)\quad U = V(\overset{1}{v}(\overset{1}{q}_1, \ldots, \overset{1}{q}_{n_1}), \ldots, \overset{j}{v}(\overset{j}{q}_1, \ldots, \overset{j}{q}_{n_j}), \ldots, \overset{m}{v}(\overset{m}{q}_1, \ldots, \overset{m}{q}_{n_m}))$$

$$j = 1, \ldots, m < n; \qquad \sum_{j=1}^{m} n_j = n$$

---

[3] Earlier, Frisch (p. 17) proved that homotheticity under the different name of "expenditure proportionality" is a sufficient condition for equation (7). Commenting on Frisch, Abram Bergson proved the necessity of this restriction, pp. 42–45, 49–50. Shephard proved the theorem again (1953, pp. 43–48 and 1970, p. 304). Homotheticity has long been known in the theory of price indices for another effect. Frisch showed that if the condition of expenditure proportionality holds, the true price index lies between an upper bound determined by the Laspeyres Index and a lower bound determined by the Paasche Index.

[4] The Hicks condition of fixed relative prices for the grouped commodities provides an alternative sufficient condition.

[5] If T is derived from the indirect utility function, homogeneous separability of ψ is a sufficient condition. The indirect utility function is homogeneously separable if and only if the direct utility function is homogeneously separable. (Similarly U is homothetic if and only if ψ is.) This is an application of the duality properties of direct and indirect utility functions (see Samuelson (1965)).

where V is continuous, twice differentiable, and strictly quasi concave and the v_i are continuous, twice differentiable functions, homogeneous of degree one.

W. M. Gorman has shown that if the utility function is homogeneously separable, commodities may be consistently aggregated in the sense that one may form composite commodities from the homogeneously separable groups which may be treated in the same manner as the primary commodities. In particular, the commodity demand functions of composite commodities so formed have all the usual properties (see, for example, Donald Katzner, pp. 140–43). One immediate implication is that the expenditure function, and hence the true price index, is separable in the same grouping of commodities.[6] Moreover, it can be shown that the price functions for the composite commodities are themselves homogeneous of degree one in the prices of the component commodities (see Shephard (1970, ch. 6)). We may therefore form ratios of the prices $(\bar{p}_j^1/\bar{p}_j^0)$ which are price sub-indices satisfying, like T, all the Fisher test criteria (see Samuelson and Swamy for discussion of these tests). T may also be written as a mean of the ratios of the prices $(\bar{p}_j^1/\bar{p}_j^0)$.

A few particular functions will illustrate these principles. First, consider the Stone-Geary utility function which has been used to estimate true indices:

$$U = \sum_{i=1}^{n} \beta_i \, log\,(q_i - \gamma_i)$$

$$0 < \beta_i < 1, \quad \sum_i \beta_i = 1, \quad q_i > \gamma_i \geq 0$$

The associated true price index may be derived by integration of the derived commodity demand functions (see Klein and Rubin, and Samuelson (1947)), or from the indirect utility function (see Goldberger), or from the expenditure function. The function e may be found to be

$$(10) \quad e = \sum_{i=1}^{n} p_i\gamma_i + c\prod_{i=1}^{n} p_i^{\beta_i}$$

$$c = c(U) = \text{constant for } U = \mu$$

In its most convenient form T may be written as

$$(11) \quad T = s^{0}\sum_{i=1}^{n} w_i^{0}(p_i^{1}/p_i^{0}) + t^{0}\prod_{i=1}^{n} (p_i^{1}/p_i^{0})^{\beta_i}$$

Thus the true index is the weighted arithmetic mean of two sub-indices as required by equation (6). The first is the sub-index of the cost of buying the subsistence quantities of the commodities $(\gamma_1, \ldots, \gamma_n)$, and the second is the sub-index of the cost of buying the above-subsistence quantities of the same commodities. These two sub-indices are themselves weighted arithmetic and geometric means of price relatives, the weights being the corresponding shares of subsistence and above-subsistence expenditure, respectively.

It may be noted that the Stone-Geary utility function is homothetic in the variables $(q_1-\gamma_1, \ldots, q_n-\gamma_n)$, that is, the above-subsistence quantities consumed of the commodities. Alternatively, the Stone-Geary function may be said to be homothetic to the point $(\gamma_1, \ldots, \gamma_n)$. This is a generalization of the definition of homotheticity to the origin. Equation (11) exemplifies that the true price index for a function which is homothetic to some point other than the origin is not independent of the initial level of income.

An example of a homothetic function is the n-commodity (Bergson) or generalized *CES* function (see Hirofumi Uzawa):

$$(12) \quad U = \left[\sum_{i=1}^{n} \alpha_i q_i^{-\rho}\right]^{-1/\rho} \qquad \alpha_i > 0, \rho > -1$$

---

[6] Shephard (1953) proved that homogeneous separability was sufficient for aggregation of the separable variables in the cost function of a producer but did not apply the result to the theory of price indices.

This function has the special property that the partial elasticity of substitution (σ) between any pair of commodities is constant and identical for all pairs. The expenditure function is (see Uzawa, p. 293):

$$(13)\quad e = U\Big[\sum \overset{\sigma}{\alpha}_i p_i^{1-\sigma}\Big]^{1/1-\sigma} \qquad \sigma = 1/(1+\rho)$$

It is easily shown that T can be written in the useful form

$$(14)\quad T = \left[\sum_{i=1}^{n} \overset{0}{a}_i(\overset{1}{p}_i/\overset{0}{p}_i)^{1-\sigma}\right]^{1/1-\sigma}$$

where

$$a_i = p_i q_i/e$$

$$= \overset{\sigma}{\alpha}_i(p_i/p_n)^{1-\sigma}\left[\sum_j \overset{\sigma}{\alpha}_j(p_j/p_n)^{1-\sigma}\right]^{-1}$$

It is seen that a_i, the share of each commodity in total expenditure, depends solely on relative prices. The true price index is independent of the initial level of utility as required by Theorem 1. Moreover, equation (14) shows that the index takes a special form. It is the weighted mean of order $(1-\sigma)$ in the price relative variables, the weights being the initial period budget shares. This mean form follows from the expenditure function being a weighted mean of order $(1-\sigma)$, and its dual the direct utility function a weighted mean of order $-\rho$ (see Kenneth Arrow, et al.). Thus the expenditure function e (and hence the true index T) and U are self-duals in the sense that they have the same functional form.[7]

As special cases with $\sigma = 0$ and 2, we have the weighted arithmetic and harmonic means, respectively. The limiting case of the *CES* function as $\sigma \rightarrow 1$ or that of the Stone-Geary function as $\gamma_i \rightarrow 0$ for all i, is the Cobb-Douglas function

$$(15)\quad U = \prod_{i=1}^{n} q_i^{\beta_i} \qquad \beta_i = \alpha_i \Big/ \sum_{i=1}^{n} \alpha_i$$

Its associated price index is, from equation (11) or (14), (see G. H. Hardy, J. F. Littlewood and G. Polya, p. 15):

$$(16)\quad T = \prod_{i=1}^{n} (\overset{1}{p}_i/\overset{0}{p}_i)^{\beta_i} \qquad \sum_{i}^{n} \beta_i = 1$$

Thus, the geometric mean is derived as an exact true price index under the hypothesis that U is Cobb-Douglas.

The use of homogeneous separability may be exemplified by the two-level generalizations of the *CES* functions which were developed in production theory. The utility function corresponding to the two-level Sato generalization is

$$(17)\quad U = \left[\sum_{j=1}^{m} \alpha_j(\bar{q}_j)^{-\rho}\right]^{-1/\rho} \qquad \alpha_j > 0 \quad \rho > -1$$

where $\bar{q}_j$ is a quantum index function of the form

$$(18)\quad \bar{q}_j = \left[\sum_i \overset{j}{\alpha}_i(\overset{j}{q}_i)^{-\rho_j}\right]^{-1/\rho_j}$$

$$\overset{j}{\alpha}_i > 0, \quad \rho_j > -1$$

Thus U in equation (17) is a *CES* function in variables $\bar{q}_j$ which, as in (18), are themselves *CES* functions of the quantities of commodities actually marketed. The first function of the composite commodities $\bar{q}_j$ in equation (17) will be described as the "top level," and the second function in equation (18) as the "bottom level." The partial elasticities of substitution between any two commodities i and k are given by (see Sato):

> $\sigma_{ik} = \sigma_j$ if i and k are in the same group j, and a harmonic mean of σ, σ_r, and σ_s if i and k are in two distinct groups, r and s.

The implied expenditure function is shown as (19).

---

[7] Self-duality of the direct and indirect utility functions U and ψ, has been considered by Samuelson (1965, p. 793n) and Houthakker.

$$(19) \qquad e = U\left[\sum_{i=1}^{m} \alpha_j^{\sigma} \bar{p}_j^{\,1-\sigma}\right]^{1/1-\sigma}$$

where $\bar{p}_j$, the price of the commodities in the jth group is

$$(20) \qquad \bar{p}_j = \left[\sum_{i=1}^{n} \alpha_i^{\sigma_j} (\overset{j}{p}_i)^{1-\sigma_j}\right]^{1/1-\sigma_j}$$

$$\sigma_j = 1/(1+\rho_j)$$

Hence, by the definition of T and applying equation (14)

$$(21) \qquad T = \left[\sum_{j=1}^{m} \overset{0}{\bar{a}}_j (\overset{1}{\bar{p}}_j/\overset{0}{\bar{p}}_j)^{1-\sigma}\right]^{1/1-\sigma}$$

$$\bar{a}_j = \bar{p}_j \bar{q}_j / e$$

where

$$(22) \quad \overset{1}{\bar{p}}_j/\overset{0}{\bar{p}}_j = \left(\sum_i [(\overset{j}{a}_i)^0][(\overset{j}{p}_i)^1/(\overset{j}{p}_i)^0]^{1-\sigma_j}\right)^{1/1-\sigma_j}$$

$$\overset{j}{a}_i = \overset{j}{p}_i \overset{j}{q}_i / \overset{j}{e}$$

T is a weighted mean of order $(1-\sigma)$ in variables which are themselves weighted means of order $(1-\sigma_j)$.

We see from equation (21) that the true price index for the two-level *CES* function is independent of the initial level of expenditure and is self-dual with the two-level direct utility function and expenditure functions. Finally, the index may be written in terms of the price relatives $(\bar{p}_j^1/\bar{p}_j^0)$. These are price sub-indices for each separable group of commodities. That T may be written in this form is the result of the joint restrictions of homogeneity and homogeneous separability. We shall refer to the index in equation (21) which is a function of the sub-indices as the top-level index and to the sub-index for each commodity group in equation (22) as the bottom-level indices.

II

The most important question is the extent of the bias that arises in empirical studies of price movements, where a non-true price index is employed when a true price index should be employed. In most cases the index employed is a Laspeyres or Paasche index with fixed base or current period weights, respectively. In all cases the extent of the bias will depend on the pattern of relative price and income movements between the situations compared and the changes in consumption in response to these price changes. In the case of fixed period weights the bias arises because these indices ignore the substitution of now relatively cheaper commodities for those whose relative prices have increased, which is a consequence of utility maximization. In the cases of other indices, the bias arises either because the index is a true price index derivable from some utility function which may not be a sufficiently good representation of consumer's preferences, or because the set of weights used is not consistent with any utility maximization process.

In any particular instance the extent of the bias upon discrete changes will of course depend on both the index which has been arbitrarily chosen and on the particular utility function which it is believed truly represents the consumer's preferences. The function selected for this purpose is the Sato two-level *CES* function with two groups of two commodities each, a total of four commodities.

It may be objected that the results below are limited by the use of a particular functional form. This function is restricted in that it is homogeneous of degree one and the intragroup elasticities of substitution for commodities σ_j and of substitution between commodity groups σ are constant. Homogeneity is not a serious additional restriction for production processes but the reality of it, or rather its homothetic transformations, may be questioned for the analysis of consumer behavior. However, in this context homogeneity is an advantage as it removes the income effect. Although the true price index is defined with respect to a chosen level of utility, the arbitrary choice of this level for a time-series may affect substantially the values of the true index and the bias from calculating a nontrue index if the utility function is not homothetic and the income elasticities differ markedly among commodities. Similarly, the constancy of the partial elasticities of substitution between commodities within the same group and between commodity groups allows one to measure the importance of substitution effects directly by varying these substitution parameters. An additional advantage of the function is that the Laspeyres, geometric, and harmonic means are all special cases of the derived price index.

Table 1 presents a range of values for the index T^* for selected representative price changes and parameter values. The index T^* is

$$T^* = 100[a_x r_x^{1-\sigma} + a_y r_y^{1-\sigma}]^{1/(1-\sigma)} \tag{23}$$

where r_x and r_y are the chosen price relatives. The expression has been multiplied by 100 to put it in the form conventional for computation. It can apply to either the top level or the bottom level. To obtain the true price index for the n commodities, one first applies this expression to obtain changes in the commodity prices for each group of commodities at the bottom level given the selected parameter values. One then repeats the calculation at the top level given the computed sub-price indices and the selected values for other top level parameters. In general, the parameter values will differ between the two levels.

The commonly used Laspeyres index will be used to show the bias from misspecification. The divergence between the true index for given price changes and the Laspeyres index with fixed base period weights *at one level* can be seen from Table 1. The use of fixed weights corresponds to the assumption of zero substitution in response to relative price changes. The Laspeyres index is an upper bound of the true index. The values of this index are given by the first row of the table with $\sigma=0$. Since errors occur at both the top and bottom levels if the utility and true price index functions are assumed to be Laspeyres when in reality they are two-level *CES* with $\sigma>0$ and $\sigma_j>0$ for $j=1$ and 2, Table 1 alone does not reveal the total bias from the use of the Laspeyres index.

TABLE 1—T^* FOR REPRESENTATIVE PRICE CHANGES[a]

	$a_x=0.3$			$a_x=0.7$		
σ	$r_y=120$	$r_y=110$	$r_y=100$	$r_y=120$	$r_y=110$	$r_y=100$
0.00	123.00	116.00	109.00	127.00	124.00	121.00
0.20	122.98	115.98	108.83	126.98	123.93	120.84
0.40	122.98	115.97	108.67	126.98	123.87	120.67
0.60	122.95	115.79	108.51	126.95	123.79	120.50
0.80	122.93	115.72	108.35	126.93	123.72	120.33
1.00	122.92	$\boxed{115.65}$	108.19	126.92	$\boxed{123.65}$	120.16
1.20	122.90	115.59	108.03	126.89	123.57	119.99
1.50	122.88	115.49	$\boxed{107.81}$	126.87	123.46	119.72
2.00	122.83	115.32	$\boxed{107.44}$	126.83	123.28	$\boxed{119.27}$
2.50	122.82	115.17	107.09	126.82	123.12	118.81
5.00	122.60	114.43	105.57	126.56	122.10	116.38

[a] It is assumed that $r_x=130$.

TABLE 2—PERCENTAGE DEVIATION OF CHANGE IN AVERAGE PRICE LEVEL FROM TRUE VALUE

	$a_1=0.3$			$a_1=0.7$		
σ	$r_1=107.81$ $r_2=119.72$	$r_1=107.44$ $r_2=123.65$	$r_1=115.65$ $r_2=123.65$	$r_1=107.81$ $r_2=119.72$	$r_1=107.44$ $r_2=123.65$	$r_1=115.65$ $r_2=123.65$
0.00	7.74	3.78	1.65	10.72	9.76	1.94
0.20	7.94	4.06	1.69	10.92	10.20	2.00
0.40	8.01	4.33	1.74	11.11	10.57	2.05
0.60	8.28	4.61	1.79	11.50	11.02	2.11
0.80	8.48	4.90	1.89	11.70	11.48	2.22
1.00	8.61	5.12	1.93	12.00	11.94	2.28
1.20	8.82	5.41	1.98	12.20	12.31	2.34
1.50	9.09	5.86	2.08	12.60	12.97	2.39
2.00	9.57	6.56	2.18	13.21	14.02	2.56

Table 2 presents the figures of the total percentage bias B which result from using the Laspeyres index in place of the true two-level *CES* index:

$$B = \left(\frac{L - T}{T - 100}\right) 100 \tag{24}$$

where

$$L = \left[\sum_{i=1}^{4} \overset{1}{p}_i \overset{0}{q}_i \Big/ \sum_{i=1}^{4} \overset{0}{p}_i \overset{0}{q}_i \right] 100$$

$$= \sum_{i=1}^{4} \overset{0}{a}_i (\overset{1}{p}_i / \overset{0}{p}_i) 100 \tag{25}$$

$$\overset{0}{a}_i = \overset{0}{p}_i \overset{0}{q}_i \Big/ \sum \overset{0}{p}_i \overset{0}{q}_i$$

and T is given by equation (21) with $j=1$, 2 and $n_j=2$, multiplied by 100. The bias examined is the bias in the *change* in the value of the index in later periods or situations from the value in the base or initial situation of 100. This is reasonable since the purpose of indices is to measure these changes from base values and it is this bias which is important in applications of index numbers. The bias has been computed by first computing the true index and the corresponding Laspeyres index and then using equation (24). The true index is computed in two stages. The first stage is the calculation of the T^* for each group of commodities from equation (23). For the second stage we take as a representative sample of the true price changes for the two groups of commodities certain pairs of the values in Table 1 which have been marked by a square. The parameter values at the bottom level which give rise to these sub-price indices are given at the columns and rows of this table. We then use T^* at the top level, using these sub-price indices. The top level substitution and distribution parameters are given at the columns and rows of Table 2.

One would expect the biases from neglecting substitution to increase with the elasticities of substitution. The purpose of this approach has been to indicate how other aspects of price movements and these elasticities affect the magnitude of these biases. The qualitative properties of the bias function in equation (24) emerge from Tables 1 and 2. The bias is an increasing function of the elasticities of intergroup substitution σ, the elasticities of intragroup substitution σ_j, the change in the relative prices of individual commodities (which is reflected in difference in the price relatives), the budget share at the bottom level of the commodity whose price index has changed least, and the budget share at the top level of the group of commodities whose price index is most overstated by use of the Laspeyres subindex. The effect on the bias of increasing the elasticities of substitution at either level can be verified from the theorem of means: a mean of order s is less than the corresponding mean of order t unless the components (here the price relatives) are all equal or $t \leq 0$ and one component is zero (see Hardy, Littlewood, and Polya, p. 26). In our expression s, $t=(1-\sigma)$. The influence of other parameters can be appreciated if we write the Laspeyres index in terms of the primary commodities as an arithmetic mean of sub-indices for commodity groups which are themselves arithmetic means. For the general case of n commodities,

$$L=\left[\sum_{i=1}^{n} \overset{1}{p}_i\overset{0}{q}_i \Big/ \sum_{i=1}^{n} \overset{0}{p}_i\overset{0}{q}_i\right]100 \tag{26}$$

$$=\sum_{j=1}^{m} \overset{0}{\bar{a}}_j(\overset{1}{\bar{p}}_j/\overset{0}{\bar{p}}_j)100$$

where

$$(\overset{1}{\bar{p}}_j/\overset{0}{\bar{p}}_j)100 = \sum_{i=1}^{n_j} (a_i^j)^0[(p_i^j)^1/(p_i^j)^0]100$$

The commodities may be grouped in the same sets as in the index for the two-level *CES* function. Substituting equation (26) for (25) and putting m and $n_j=2$ shows the influence of the weighting pattern.

Table 2 shows that the magnitude of biases due to the neglect of commodity substitution in the calculation of the price index may be significant. The values of the elasticities of substitution within and between groups are quite moderate; σ_j was assumed to take values of 1.00, 1.50, and 2.00, and σ was assumed to take values between 0 and 2.00. The intragroup elasticity of substitution in particular may reasonably take on values higher than those chosen. The chosen values of the price relatives also reflect quite small changes in the relative prices of commodities; for example, the price relatives of $r_x=130$ and $r_y=110$ imply that the price of x relative to the price of y has risen by 18 percent. Although a bias of 10 percent may not appear high absolutely, it means, for example, that an increase in the price level as measured by a Laspeyres index of 5 percent is in reality only 4.55 percent. Changes of this magnitude would have an important influence on wage, welfare payments, and other policies which are based on cost-of-living adjustments. Other realistic combinations of parameter values can give considerably greater or smaller biases than those reported here.

It is possible to decompose the figures of total bias in Table 2 into the two components, that which is due to the neglect of substitution among individual commodities at the lower level, and that which is due to the neglect of substitution among the commodity groups at the higher level. The first row of Table 2 shows the bias which results when there is substitution at the bottom level which is wrongly neglected but no substitution at the top level. The lower rows in the table show the bias when substitution at both levels is neglected. The differences between a figure in a lower row and the figure in the first row for the same column is the net contribution to the total bias of the neglect of top-level substitution.

One notable result emerges for all columns in Table 2. In all cases the neglect of substitution at the bottom level causes more than one-half of the total bias. This arises for two reasons. First, the values chosen of the elasticities of substitution among individual commodities at the bottom level are generally higher than those of the intergroup elasticity of substitution at the top level; the three values of σ_j are 1.00, 1.50, and 2.00, while the values of σ in Table 2 range from 0 to 2.00. Second, the formation of the sub-price indices at the bottom level as weighted averages of the component price changes for individual prices damps the relative price changes among the commodity groups, for a given set of price changes for the individual commodities. Thus studies which use aggregated price and quantity data to calculate price indices may well omit the most important source of bias even if they employ true price index forms.

These results suggest strongly that the two previous empirical studies of true cost-of-living indices by Goldberger and Gamaletsos for Greece, and Tran Van Hoa for Australia understated the bias due to the use of Laspeyres indices. First, as equation (11) reveals, the true index on the assumption that the particular form of the utility function is Stone-Geary differs from the Laspeyres index only with regard to above-subsistence expenditure. In the case of Greece the share of subsistence expenditure in the total budget s^0 used in the computation of T was 82 percent, so that it is not surprising that the study showed negligible bias (see Goldberger and Gamaletsos, p. 395); in the above sense, less than 1 percent. In the case of Australia the share of subsistence expenditure is 59 percent (see Tran Van Hoa, Table II). The bias in the above sense for the all-Australia index works out to be 7.63 percent. Comparing the Stone-Geary index with the two-level *CES* index, the Stone-Geary assumption that part of the budget must be devoted to an unchanging bundle of subsistence commodities rules out substitution bias among these purchases. However, the more important reason for believing that the bias may well be greater than indicated in these two studies is that both employed highly aggregative consumption data; five commodity groups in the Goldberger-Gamaletsos study and nine in the Tran Van Hoa study. Even if one accepts the maintained hypothesis that the utility function is Stone-Geary in these commodity groups, there may be substantial substitution among the individual food and clothing, etc. commodities within these groups. The neglect of these substitutions may mean most of the bias persists in the "true" estimates.

The same computations of bias may be made when an index other than Laspeyres is used. When a Paasche index with fixed current period weights is used in place of a true two-level mean, the qualitative behavior and magnitude of the bias may be determined from the analysis of the Laspeyres index bias. This is apparent if we rewrite the Paasche index as a weighted arithmetic mean of arithmetic means as was done for the Laspeyres index in equation (26), and rewrite the true index in terms of current rather than base period budget shares:

$$(27)\quad P = \left[\sum_{i=1}^{n} \overset{1}{p}_i \overset{1}{q}_i \Big/ \sum_{i=1} \overset{0}{p}_i \overset{1}{q}_i \right] 100$$

$$= \left[\sum_{j=1}^{m} \overset{1}{\bar{a}}_j [(\overset{1}{\bar{p}}_j / \overset{0}{\bar{p}}_j)^{-1} 100] \right]^{-1}$$

and

$$(28)\quad T = \left[\sum_{j=1}^{m} \overset{1}{\bar{a}}_j [(\overset{1}{\bar{p}}_j / \overset{0}{\bar{p}}_j) 100]^{-(1-\sigma)} \right]^{-1/(1-\sigma)}$$

At the bottom level the sub-price indices for the Paasche and true index are weighted arithmetic means and means of order $(1-\sigma_j)$, respectively. Given the current periods weights the values of P and T for given values at $(1-\sigma)$ and $(1-\sigma_j)$ may be identified with the values of T^* in Table 1 for values of $-(1-\sigma)$ and $-(1-\sigma_j)$.

What are the implications of these results for index calculation? First, it appears that the extent of the biases resulting from the use of fixed weight indices are too great to be ignored when substitutability among commodities is not low. The problem is likely to be acute when the base weights used for the index are not revised over a period of several years and relative prices diverge substantially from those of the base period.

Unfortunately, this is often the case. More frequent revisions of the weights and chaining of the resulting indices might reduce or increase the Laspeyres index (see for example Allen, pp. 387–88), but cannot give a true index for a chosen level or levels of utility. The predominance of the bias due to the neglect of substitution within commodity groups suggests we may obtain a good approximation to the true index in some cases if we use Laspeyres indices for the commodity groups and their prices, provided we calculate the sub-price indices correctly with the non-Laspeyres indices. Another possibility is the use of some other average; for example, the weighted geometric mean. The use of the geometric mean is equivalent to restricting all elasticities of substitution to unity but, in the presence of cross-substitution effects, it frequently gives estimates of the true value which are closer than restricting all to the value of zero by using constant weights (see the rows for $\sigma=1$ in Tables 1 and 2). However, if the degree of substitutability does differ among individual commodities or commodity groups a direct estimation of the true indices is preferable.

III

The use of these methods to examine true price indices for consumption commodities can be extended to incorporate leisure, multiperiod consumption, and durable goods.

When leisure is added as another commodity whose price is the wage rate per period of leisure foregone, income received from work and expenditure on commodities other than leisure are now endogenously determined. Although minimum expenditure now includes expenditure on the commodity leisure, the utility-maximizing definition of the true price index is no different from that of the previous sections.

The generalization of the model to allow for multiperiod consumption is quite straightforward if one uses the standard Fisherian model with no uncertainty regarding prices, incomes, and interest rates in each period and a fixed planning period. This model permits borrowing and saving in individual periods. The consumption of a physically identical commodity in each time period (or location) is regarded as distinct commodities, all of which enter the intertemporal utility function separately as arguments. These include durable goods with the (discounted) period rental of the commodity as the price.[8] Maximization of utility, given the prices in all periods, yields a time-path of consumption bundles and hence of expenditures in each period. For the purpose of the true price index, expenditure in a price situation is now defined as the aggregate present discounted value of this stream of expenditures and the true index is the ratio of minimum discounted expenditures for two chosen price situations.[9] All of the theorems above continue to apply.

Theorem 2 can be applied in the multiperiod model to achieve a further simplification if one is prepared to assume that U is homogeneously separable with respect to the period bundles, viz.,

---

[8] When the commodity is not rented, the quantity purchased and quantity available for use (equal to the depreciated stock plus new purchases) are no longer equal in each period. The model may include this case (see M. J. Martin).

[9] This definition differs from that of Malcolm Galatin. His two-period model departs from the standard model in that he assumes a consumer's plans for period 0 are fulfilled but that in period 1 there is an unexpected change in prices. The denominator of his true index is actual expenditure in period 1 and the numerator is the expenditure that would be required in period 1, given new prices in this period, to bring the consumer to the level of utility he would have achieved had the prices of period 0 continued in period 1. Since the prices of period 0 are neither the expected nor actual prices of period 1, this is a curious definition. The definition above is the logical extension of the single period case.

$$(29)\quad U = V(v^0(q_1^0, \ldots, q_n^0), \ldots, v^t(q_1^t, \ldots, q_n^t))$$

where the v^i are all homogeneous of degree one. The superscripts indicate the time periods of each consumption bundle. In this case consumption in each period can be treated as a composite commodity which has a price index. The allocation of aggregated discounted income among periods now depends only on relative period price indices. For example, the two-level Sato function of Section II can be interpreted as a two-period consumption model with the allocation of expenditures between the two periods at the top level being determined by period price indices. The numerical results of Section II apply immediately with the minimum expenditures and the commodity prices of both situations interpreted as discounted.

The theory of true price indices has applications in all areas of optimizing behavior. In production theory there is a true price index showing the average change in the cost of producing some commodity or fixed bundle of commodities. In some countries there are separate indices of materials cost and the cost of capital equipment. This separation of the current and capital costs of production might be justified by assuming the production function is homogeneously separable in these two groups of inputs. But in this case, as in the computation of almost all indices of prices, there are biases due to the neglect of substitution among commodities.

REFERENCES

R. G. D. Allen, "Price Index Numbers," *Rev. Int. Statist. Inst.*, 1963, *31*, 281–301.

K. J. Arrow et al., "Capital-Labor Substitution and Economic Efficiency," *Rev. Econ. Statist.*, Aug. 1961, *43*, 225–50.

A. Bergson, "Real Income, Expenditure Proportionality, and Frisch's 'New Methods of Measuring Marginal Utility'," *Rev. Econ. Stud.*, Oct. 1936, *4*, 33–52.

F. M. Fisher and K. Shell, *The Economic Theory of Price Indices*, New York 1972.

R. Frisch, "Annual Survey of General Economic Theory; The Problems of Index Numbers," *Econometrica*, Jan. 1936, *4*, 1–38.

M. Galatin, "A True Price Index When the Consumer Saves," *Amer. Econ. Rev.*, Mar. 1973, *63*, 184–94.

R. C. Geary, "A Note on 'A Constant-Utility Index of the Cost of Living'," *Rev. Econ. Stud.*, 1950, *18*, 65–66.

A. S. Goldberger, "Functional Form and Utility: a Review of Consumer Demand Theory," workshop pap. no. 6703, Soc. Syst. Res. Inst., Univ. Wisconsin 1967.

——— **and T. Gamaletsos**, "A Cross-Country Comparison of Consumer Expenditure Patterns," *European Econ. Rev.*, Spring 1970, *1*, 357–99.

W. M. Gorman, "Separable Utility and Aggregation," *Econometrica*, July 1959, *27*, 469–81.

G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities*, Cambridge 1952.

J. Tran Van Hoa, "Additive Preferences and Cost-of-Living Indexes: an Empirical Study of Australian Consumers' Welfare," *Econ. Rec.*, Sept. 1969, *45*, 432–40.

H. S. Houthakkar, "Self-Dual Preferences," *Econometrica*, Oct. 1965, *33*, 797–801.

D. W. Katzner, *Static Demand Theory*, London 1970.

L. R. Klein and H. Rubin, "A Constant-Utility Index of the Cost of Living," *Rev. Econ. Stud.*, 1948, *15*, 84–87.

L. McKenzie, "Demand Theory Without a Utility Index," *Rev. Econ. Stud.*, June 1957, *24*, 185–89.

M. J. Martin, "Durable Goods as a Generalizer of Demand Theory," in M. Morishima et al., eds., *Theory of Demand—Real and Monetary*, Oxford 1973, 184–200.

P. Samuelson, *Foundations of Economic Analysis*, Cambridge, Mass. 1947.

———, "Some Implications of 'Linearity'," *Rev. Econ. Stud.*, 1948, *15*, 88–90.

———, "Using Full Duality to Show that Simultaneously Additive Direct and Indirect Utilities Implies Unitary Price Elasticity of Demand," *Econometrica*, Oct. 1965, *33*, 781–96.

——— and S. Swamy, "Invariant Economic Index Numbers and Canonical Duality: Survey and Synthesis," *Amer. Econ. Rev.*, Sept. 1974, *64*, 566–93.

K. Sato, "A Two-Level Constant-Elasticity-of-Substitution Production Function," *Rev. Econ. Stud.*, Apr. 1968, *34*, 201–18.

D. W. Shephard, *Cost and Production Functions*, Princeton 1953.

———, *Theory of Cost and Production Functions*, Princeton 1970.

H. Theil, "The Information Approach to Demand Analysis," *Econometrica*, Jan. 1965, *33*, 67–87.

H. Uzawa, "Production Functions with Constant Elasticities of Substitution," *Rev. Econ. Stud.*, Oct. 1962, *29*, 291–99.

A Labor Theoretic Analysis of the Criminal Choice

By M. K. Block and J. M. Heineke*

> "Much of the security of person and property in modern nations is the effect of manners and opinion rather than law."
>
> John Stuart Mill
> *Principles of Political Economy*

Recently, a number of economists have applied modern choice theory to the study of illegal or criminal activities. Almost without exception, they have emphasized the similarity between the decision to commit an offense and the traditional household choice problem. As Gary Becker in his pioneering article expresses the proposition: "Some persons become 'criminals' . . . not because their basic motivation differs from that of other persons, but because their benefits and costs differ" (p. 176). Although this point is well taken, we find that a number of recent contributions do not provide an adequate framework for analyzing the costs and benefits of an important class of illegal activities. In particular, Becker, Isaac Ehrlich, and David Sjoquist summarize the consequences of time-consuming illegal activities in terms of a distribution on wealth alone without fully considering the underlying multiattribute choice problem.[1]

The commission of most offenses results in an expenditure of effort, the possibility of an increase in the individual's wealth position, and the possibility of a penalty. Aside from the penalty, the similarity between such offense decisions and labor supply decisions under uncertainty is obvious.[2] Moreover, if the penalty is a monetary payment, the analogy is precise.

Although many criminal choice problems may be viewed within an expanded labor choice framework, care must be exercised if these problems are to be interpreted in terms of strictly monetary costs and benefits. We show below that by not fully specifying their choice problems, and therefore the transformation between what is inherently a multiattribute decision problem and the wealth-only problem, Becker, Ehrlich, and Sjoquist are led to conclusions which are valid only in very special cases. In general, we show that plausible preference restrictions are not sufficient to generate unambiguous supply results, a result that should come as no surprise since it is the same situation that confronts the investigator in most household allocation problems.[3] Therefore, policy prescriptions in this area, as in the tax incentive area, do not follow from theory but rather require empirical determination of relative magnitudes.

---

* Associate professors of economics, Naval Postgraduate School and University of Santa Clara, respectively. We are indebted to James Sweeney, Hayne Leland, Agnar Sandmo, and Henry Demmert for their comments and suggestions. In addition, we would like to acknowledge the comments of a referee which have been most helpful in clarifying several portions of our argument. An earlier version of this paper was read at the European Meetings of the Econometric Society 1973.

[1] Michael Allingham and Agnar Sandmo, Serge-Christophe Kolm, and Balbir Singh also have papers on the criminal choice problem. But each of these papers is concerned with income tax evasion, an activity in which "labor" may be a relatively insignificant input. If this is the case, modeling the decision problem as a choice over wealth orderings is appropriate. However, for the criminal choice in general, the labor attribute will be significant and must be included in the agent's preference orderings. For this reason, most of our attention will be focused on the papers of Becker, Ehrlich, and Sjoquist who model the general criminal choice problem. Allingham and Sandmo are aware that several attributes must in general be included in the individual's decision problem and examine certain aspects of this problem in one section of their paper.

[2] See the authors (1973a) for an analysis of the labor supply decision when returns are stochastic.

We proceed as follows: First, the individual's labor-offense supply problem is formulated in terms of the underlying multiattributed nature of the problem. Next, supply responses to various parameter shifts are investigated. As would be expected, under "usual" preference restrictions these responses are ambiguous in sign. Finally, since unambiguous results have been reported in the literature, the last section and the Appendix are devoted to determining the conditions under which unequivocal supply effects obtain.

I. The Joint Supply of Labor and Offenses

In the analysis which follows we restrict our attention to property crimes, which enables us to concentrate on what George Stigler refers to as "production offenses." Specifically, we consider an individual who is confronted with two wealth generating activities, legal activity (labor) and illegal activity (theft) and denote the time spent in these activities as L and T, respectively. Hence, the individual's evaluation of his well-being at any point in time will be a function of the time spent generating wealth and the level of his wealth, i.e.,

$$U = U(\mathrm{L}, \mathrm{T}, W) \tag{1}$$

where U is the agent's von Neumann-Morgenstern utility indicator, and W represents wealth, with $U_W>0$, $U_{\mathrm{L}}<0$, and $U_{\mathrm{T}}<0$. By including the arguments L and T explicitly in U, we are provided with a straightforward means of analyzing the role of moral and ethical considerations which may constrain the work-theft decision.

The following definitions will be used:

$r \equiv$ the rate of return to legal activity
$V \equiv$ the rate of return to illegal activity
$a \equiv$ the stochastic failure, capture, or arrest rate, $0 \leq a \leq 1$
$\theta \equiv$ the number of offenses, $\theta=\theta(\mathrm{T})$ and $\theta'(\mathrm{T})>0$
$F \equiv$ the fine per offense
$\hat{W} \equiv W^0 + r\mathrm{L} + (V-aF)\theta(\mathrm{T})$, actual wealth[4]
$\mathrm{N} \equiv$ time devoted to nonmarket activity
$\bar{\mathrm{t}} \equiv \mathrm{L}+\mathrm{T}+\mathrm{N}$

Note that the penalty for an offense is specified as a fine. This penalty specification enables us to focus on an issue of central concern in this paper, the role of psychic costs in the offense decision.[5]

A. *The Model*

According to the expected utility theorem, the individual's labor-theft supply decision is determined by

$$\max_{\mathrm{L},\mathrm{T}} \int U[\mathrm{L}, \mathrm{T}, W^0 + r\mathrm{L} + (V - aF)\theta]f(a)da \tag{2}$$

subject to the condition that labor and theft levels be nonnegative. In (2), $f(a)$ is the agent's subjective probability density on the arrest rate and indicates the agent's beliefs as to the intervals in which the arrest rate is likely to lie. To facilitate comparison to the existing literature, we adopt the specification used in Ehrlich, Sjoquist, and Morgan Reynolds and fix the amount of time allocated to nonmarket activities.[6] Further, the number of offenses is assumed to be proportional to the time devoted to their production. Under these assumptions the first-order condition for a relative maxima requires that

---

[3] In particular, the agent's simple behavior toward risk contains qualitative supply implications only in highly restrictive circumstances.

[4] We use the term "actual wealth" to denote the wealth that an individual has available to meet financial obligations. It is initial wealth W^0 plus earnings or losses during the period under consideration; $\hat{W}$ is a particular value of W.

[5] While the introduction of a prison sentence would complicate the analysis, it would not invalidate the basic argument which follows. In addition, according to Becker, pp. 193–98, fines are not only the most common form of punishment, but also the most "efficient." For an analysis in which prison sentences are formally introduced into the choice problem, see Block and Robert Lind.

$$E[U_T - U_L + U_W((V - aF)\theta' - r)] \leq 0 \tag{3}$$

where $\theta' \equiv d\theta/d$T. As would be expected, when the psychic cost of effort is afforded its traditional labor theoretic role, the agent's simple behavior toward risk (*sign* U_{WW}) has no unique allocative implications.[7] Hence, Ehrlich's assertion, p. 528, that preferences toward risk and relative returns alone determine the degree of specialization will not hold in general. Only in a special case where "returns" include a strong assumption concerning psychic costs is such a statement valid.[8] In general, the time allocation between L and T will depend not only upon the agent's behavior toward risk and relative returns but also upon the relative "irksomeness" of alternative occupations.

By way of illustration, consider an individual for which $U_L - U_T > 0$ for all L, T, and W. We might say such an individual has a *preference for honesty*. If he is also risk averse, then a necessary but not sufficient condition for T>0 is that the returns to illegal activity be greater than expected costs (where costs consist of the average penalty plus legal opportunities foregone). For this condition to also be sufficient for T>0, returns must be sufficiently high to outweigh the psychic disadvantage of participation in illegal acts. In addition, increasing the certainty of arrest, increasing penalties, or increasing legal opportunities until "crime does not pay," $(V - E(a)F)\theta' - r < 0$, will deter this group of offenders. On the other hand, if the individual displays both a preference for risk and honesty, making "crime not pay" may not deter participation.[9]

II. Supply Behavior and Policy Changes[10]

In this section, we pose a number of questions concerning the supply behavior of a single economic agent. In particular, we investigate the agent's supply response to changes in (i) initial wealth, (ii) the payoff to illegal activity, (iii) the arrest rate, and (iv) the severity of punishment.

In most of the comparative static derivatives which follow, second derivatives of U appear in product expectations. This points up a well-known characteristic of stochastic models, viz., that a qualitative analysis of parameter shifts in these models often requires third derivative information concerning the agent's utility indicator.[11] The customary method of providing this information is to postulate plausible hypotheses regarding the agent's behavior toward risk as various arguments of the utility indicator change. For example, the multiattribute analog of the Arrow-Pratt absolute risk aversion function, $R \equiv -U_{WW}/U_W$, has been termed the conditional (absolute) risk aversion for wealth by Ralph Keeney. If the agent becomes increasingly willing to accept a wealth gamble of a given size as his wealth increases, *ceteris paribus*, he is said to display decreasing absolute risk aversion in wealth ($\partial R/\partial W<0$), a hypothesis we shall adopt. This restriction on the agent's preferences has been widely utilized and has led to many interesting results.[12]

---

[6] That is, $\mathrm{N} = \bar{\mathrm{N}}$, a constant.

[7] Notice that although only first derivatives of U appear in (3), a necessary condition for signing the term $EU_W(V-aF)$ is knowledge of *sign*$[U_{WW}]$. To see this note that $EU_W(V-aF) \equiv \mathrm{Cov}(U_W, V-aF) + E(U_W)E(V-aF)$ and that *sign*$[\mathrm{Cov}(U_W, V-aF)]$ depends upon how U_W *changes* with *changes* in $V-aF$ (on the average). To illustrate, consider an increase in the value of a (a decrease in $V-aF$). Now decreases in $V-aF$ cause decreases in $\hat{W}$ which in turn cause U_W to either increase or decrease depending upon whether U_{WW} is negative or positive. So if $U_{WW}<0$ then decreases in $V-aF$ cause increases in U_W and U_W and $V-aF$ move in opposite directions on the average, i.e., $\mathrm{Cov}(U_W, V-aF)<0$. Similar arguments show $U_{WW}>0$ implies $\mathrm{Cov}(U_W, V-aF)>0$ and $U_{WW}=0$ implies $\mathrm{Cov}(U_W, V-aF)=0$. In general a necessary but not sufficient condition for signing product expectations containing unrestricted random variables is sign knowledge on derivatives of one higher order than those appearing in the expectation.

[8] This point is discussed in more detail below.

[9] Risk-averse individuals with a preference for illegal activities, $U_L - U_T < 0$, may not be deterred by making crime not pay in this sense.

[10] In the discussion that follows, we assume internal solutions exist to first-order conditions.

[11] This statement is an application of the principle stated at the end of fn. 7. So, for example, a necessary condition for signing the term $EU_{WW}(V-aF) \equiv \mathrm{Cov}(U_{WW}, V-aF) + E(U_{WW})E(V-aF)$ which appears in equation (4), is *sign*$[\partial U_{WW}/\partial(V-aF)]$. That is, *sign*$[\mathrm{Cov}(U_{WW}, V-aF)]$ depends upon how U_{WW} *changes* as $V-aF$ *changes*. Hence, *sign*$[U_{WWW}]$ is needed.

A. *Wealth Effects*

A question of considerable interest to both criminologists and economists is the effect on the level of criminal activity of changes in the offender's "initial wealth." For example, would increased welfare payments have incentive or disincentive effects on the supply of offenses? To investigate this question, differentiate (3) with respect to W^0. In which case

$$\partial \mathrm{T}/\partial W^0 = E[U_{\mathrm{L}W} - U_{\mathrm{T}W} - U_{WW}((V - aF)\theta' - r)]/F_{\mathrm{TT}} \tag{4}$$

where $F \equiv EU(\mathrm{L}, \mathrm{T}, \hat{W})$.

Clearly, knowledge of the individual's simple behavior toward risk (*sign* U_{WW}) will not provide sufficient information to deduce the inferiority (or perhaps normality) of illegal activity; nor for that matter, will the combination of, say, risk aversion and decreasing absolute risk aversion. Only a priori considerations can *sign* $\partial \mathrm{T}/\partial W^0$ at this level of generality.

B. *Payoff Effects*

To our knowledge, most of the research on illegal activities has focused directly on deterrence, and hence payoff effects on the supply of these activities have been largely ignored.[13] This neglect appears even in much of the recent economics of crime literature. For example, although Becker includes "net returns" in his formulation, it is not central to his supply of offenses analysis. Certainly, any analysis of property crimes must include an examination of payoff effects as a matter of central concern. To this end, write

$$\partial \mathrm{T}/\partial V = -EU_W\theta'/F_{\mathrm{TT}} + \theta\partial \mathrm{T}/\partial W^0 \tag{5}$$

Equation (5) is the stochastic analog of the familiar Slutsky expression and is composed of a substitution effect and a wealth effect. Since F_{TT} and θ' are negative and positive, respectively, the substitution term is positive. Hence, the direction of the supply response will depend upon the wealth effect. If theft is an inferior activity, no qualitative conclusions are forthcoming.[14]

Of course, this comes as no surprise. Economists have long known that "price effects" in household decision models are ambiguous in sign. Without further preference information, the necessary condition for a positive supply response is the normality (or wealth independence) of illegal activity. Without this condition, the possibility that theft is a Giffen activity cannot be dismissed.

C. *Enforcement Effects*

In the model being investigated in this paper, uncertainty is introduced through the enforcement variable a. The payoff and penalty are both assumed to be known but the frequency of penalty imposition (the arrest rate) a is taken by the agent to be a continuous random variable, $0 \leq a \leq 1$. This specification is a generalization of the

---

[12] See, for example, Sandmo (1970, 1971), Hayne Leland (1968, 1972), Jan Mossin (1968a,b), and Block and Heineke (1972, 1973a).

[13] See Clarence Schrag, pp. 20–113, for a brief survey of the criminology literature in this area.

[14] Of course this statement remains valid a fortiori if the time allocation to nonmarket activities is endogenously determined.

Bernoulli formulation used by Becker, Ehrlich, and Sjoquist. (See the Appendix.)

The relation between the offense decision and changes in the degree of enforcement has been a topic of long-standing speculation. But because the arrest rate is a random variable there is no unique interpretation of an increase in enforcement. However, an intuitive approach is to consider changes in enforcement procedures that increase the expected number of arrests but leave all other moments of $f(a)$ unaltered. That is, we consider a "pure" increase in the arrest rate. This may be accomplished by replacing a in (3) with $a+\delta$ where δ is a mean altering, dispersion preserving parameter. Differentiating with respect to δ and evaluating at $\delta=0$ yields:

$$\partial \mathrm{T}/\partial\delta = -F(\partial \mathrm{T}/\partial V) \tag{6}$$

As we have noted, without relative magnitude information, $\partial \mathrm{T}/\partial V$ is unambiguously signed only if $\partial \mathrm{T}/\partial W^0 \geq 0$. Hence, for this class of penalties, we are able to assert unequivocally the deterrent effect of increases in the arrest rate ($\partial \mathrm{T}/\partial\delta < 0$) only by assuming the normality (or wealth independence) of illegal activity.

D. *Penalty Effects*

In the past decade we have witnessed a heated polemic concerning the effects of changes in the severity of punishment on the crime rate. Protagonists of the "liberal" position have often claimed that increasing the severity of punishment has little or no deterrent effect on the supply of offenses, while more "conservative" individuals have denounced this group as "soft on crime" and recommended increased penalties to combat growing crime rates. Although much of this argument has been couched in ideological considerations, the central question concerning the supply effects of changes in the severity of punishment is a major concern of policy makers. We now consider this question in the context of the present model.

The first-order conditions (3) indicate that the net rate of return to theft, the individual's behavior toward risk, and his "ethics," jointly determine the offense level. Hence, by examining (3) one can find several combinations of ethics and behavior toward risk which would result in zero offenses for sufficiently severe penalties. For example, if the world were comprised of risk-averse individuals who display honesty preference, $(U_{\mathrm{L}}-U_{\mathrm{T}})>0$, then the supply of offenses could be driven to zero by making F sufficiently large.[15] However, this is not likely to be possible, and if not, the question of the supply response to a change in the severity of the penalty must be formulated in terms of marginal changes in the penalty.

Since F is deterministic in the present model, the interpretation of a change in the penalty is straightforward. In fact, increases in F act as scale changes on the random variable a, decreasing expected returns and increasing the dispersion of returns. Formally,

$$\begin{aligned} \partial \mathrm{T}/\partial F = {} & E(U_W a)\theta'/F_{\mathrm{TT}} \\ & + \theta E\{a[U_{\mathrm{T}W} - U_{\mathrm{L}W} + U_{WW} \\ & \cdot((V - aF)\theta' - r)]\}/F_{\mathrm{TT}} \end{aligned} \tag{7}$$

Inspection of (7) reveals the substitution effect of a change in penalty to be negative and the wealth effect to be unsigned without further preference information. Hence, at least at the present level of generality, arguments alleging the disincentive effects of increases in the severity of punishment are not unambiguously supported by theory.

We have seen that if the multiattributed nature of the individual's decision problem is fully accounted for, then the "usual" preference restrictions concerning the individual's behavior toward risk will not provide sufficient information to sign the supply effects of increased "payoffs," "enforcement," and "penalties." The core of the problem is of course the fact that wealth effects are unsigned. And, assuming theft to be an inferior activity does not alleviate the ambiguity, since relative magnitude difficulties then arise in each case.

---

[15] See Block and Lind for a discussion of the limits of the criminal sanction.

III. Ethical Costs and Wealth: The Case of Independence

Up to this point, we have analyzed the offense decision as a generalized labor supply problem. As we have seen, the price of this generality is qualitative ambiguity.[16] In particular, the unambiguous results reported by Becker, Ehrlich, and Sjoquist are not forthcoming when the offense decision is analyzed as a general multiattribute decision problem.[17] An interesting question thus arises. What assumptions concerning the agent's utility function are implicit in the several unambiguous results reported by these authors? Or more generally, given the supply problem posed in (2), under what conditions do changes in the various components of the return to illegal activity lead to unambiguous supply responses? It is to this question that we now turn.

Becker, Ehrlich, and Sjoquist have analyzed the criminal choice using a special form of (2) in which all costs and benefits associated with criminal activity have been expressed in terms of wealth alone. That is, in their models L and T do not enter the utility function directly as attributes, but rather affect the level of utility indirectly through their effects on wealth. For example, Becker writes, costs ". . . can be made comparable by converting them into their monetary equivalent . . ." (p. 179), while Ehrlich defines the individual's wealth so that it includes ". . . assets, earnings within the period and the 'real wealth' equivalent of nonpecuniary returns from legitimate and illegitimate activity . . ." (p. 525). Or, in the words of Sjoquist: "The psychic gain is measured by that quantity of money which the individual is willing to pay to obtain the psychic gain" (p. 439). To contrast the present model with the work of these authors, we reformulate the above problem and express the psychic cost of L and T in terms of their wealth equivalents.

Formally, the problem posed in (2) may be reduced to an equivalent single attribute problem by defining a level of wealth W^* such that $U(\mathrm{L}, \mathrm{T}, \hat{W}) \equiv U(0, 0, W^*)$. Clearly, $\hat{W} - W^*$ is the wealth equivalent of L hours of legal activity and T hours of illegal activity.[18] In general, the wealth equivalent will be a function, say C, of L, T, and $\hat{W}$; i.e.,[19]

$$\hat{W} - W^* = C(\mathrm{L}, \mathrm{T}, \hat{W}) \tag{8}$$

Using (8), we may write

$$U(\mathrm{L}, \mathrm{T}, \hat{W}) \equiv U(0, 0, \hat{W} - C(\mathrm{L}, \mathrm{T}, \hat{W})) \tag{9}$$

---

[16] Again, this ambiguity persists a fortiori if the time allocated to nonmarket activities is not fixed.

[17] The only exception to this statement is the response of offenses to a change in a when $f(a)$ is Bernoulli, a result reported by all three authors. Although this result does hold after the multiattributed structure of the problem is incorporated into the model, it holds only because the arrest rate is assumed to be Bernoulli distributed. Some of the implications of assuming $f(a)$ to be Bernoulli are discussed in the Appendix.

[18] A case in which the wealth equivalent of T hours of illegal activity does not exist occurs when $(\partial W/\partial \mathrm{T})_{U^0}$ fails to exist for all L, T, and W. An individual possessing such an ethic might be said to be *absolutely honest*. In this case, one has a family of utility indicators parametric on T which are lexicographically ordered by T. Formally, let $V^{\alpha}(\mathrm{T}^{\alpha}, \mathrm{L}, W)$ represent a family of utility indicators parametric on T^{α}, where $\alpha \epsilon \Lambda$, an index set. Then absolute honesty implies $V^{\alpha_0} > V^{\alpha^*}$ iff $\mathrm{T}^{\alpha_0} < \mathrm{T}^{\alpha^*}$, for all L and W. The set $\{V^{\alpha}\}$, $\alpha \epsilon \Lambda$, is only partially lexicographically ordered since if $\mathrm{T}^{\alpha_0} = \mathrm{T}^{\alpha^*}$, then $V^{\alpha_0} \gtrless V^{\alpha^*}$ depending upon the values of L and W.

[19] A unique wealth equivalent exists (and hence the function C) iff U_{L}, U_{T}, and U_W are continuous, monotonic functions of their arguments and $U_W > 0$ everywhere.

One can analyze the choice problem at hand in terms of either the right-hand or left-hand side of (9). On the right-hand side the attributes L and T have been collapsed into their wealth equivalents, leaving the single attribute utility indicator, $U(0, 0, W^*)$. For brevity, we define $U(0, 0, W^*) \equiv \bar{U}(W^*)$. Of course, nothing has been changed since the individual's orderings over the single attribute $\hat{W} - C(\mathrm{L}, \mathrm{T}, \hat{W})$ are equivalent to his orderings over the attributes L, T, $\hat{W}$.[20] Note that the single attribute formulation treats psychic costs (benefits) as a simple subtraction (addition) from (to) wealth and hence, is the analytic justification for the approach adopted by Becker, Ehrlich, and Sjoquist in which psychic costs and returns are reduced to their monetary equivalent and then combined with monetary returns and costs. Unfortunately, none of these authors has derived his model from the underlying multiattributed structure of preferences, with the consequence that the results reported in each paper are valid only for a special case. In terms of the utility indicator $\bar{U}(\hat{W} - C(\mathrm{L}, \mathrm{T}, \hat{W}))$, their special case is equivalent to assuming the function C is independent of wealth, i.e., $C = C(\mathrm{L}, \mathrm{T})$ or $C_W = 0$.[21] We now turn our attention to several implications of this assumption.

A. *Wealth Effects* ($C_W = 0$)

As we have seen, traditional restrictions on preference orderings are insufficient to establish the effect of changes in initial wealth on the allocation of time to criminal activities. Hence the supply effects of changes in the payoff to illegal activity, the degree of enforcement, and the severity of punishment are unsigned. However, if attention is focused on the special case analyzed by Becker, Ehrlich, and Sjoquist, in which the monetary equivalent of the psychic cost of legal and illegal activity C is independent of wealth, then the wealth effect is signed and likewise are the effects of changes in the payoff, enforcement, and punishment. To see this we reformulate the supply problem given in (2) by replacing $U(\mathrm{L}, \mathrm{T}, \hat{W})$ with $\bar{U}(\hat{W} - C(\mathrm{L}, \mathrm{T}))$.[22] In which case equation (4) becomes

$$(4') \quad \partial \mathrm{T}/\partial W^0 = -E\bar{U}_{WW}[(V - aF)\theta' - r + C_{\mathrm{L}} - C_{\mathrm{T}}]/H_{\mathrm{TT}}$$

where $H \equiv E\bar{U}(\hat{W} - C(\mathrm{L}, \mathrm{T}))$.

As is obvious, the agent's simple behavior toward risk (*sign* $\bar{U}_{WW}$) provides sufficient information for signing (4′) only in the trivial case of risk neutrality, in which case the individual's time allocation to theft is invariant to changes in W^0. Generally, third derivative information will be needed.[23] If the individual is risk averse, the Arrow-Pratt measure provides the needed information. It can be shown that if this measure decreases in wealth ($\partial R/\partial W < 0$), then the numerator of (4′) is negative.[24] The crucial requirement, which is absent in the general case where $C = C(\mathrm{L}, \mathrm{T}, \hat{W})$, is that the non-linear portion of the wealth constraint be nonrandom. This is precisely the effect of making C independent of wealth. We now have

$$(10) \qquad \partial \mathrm{T}/\partial W^0 > 0$$

If the pyschic costs of effort are independent of wealth, and if the agent exhibits decreasing absolute risk aversion, then effort expended generating income via illegal activity will increase with wealth. In other words, given the widely employed and currently unrefuted hypothesis of decreasing absolute risk aversion, the Becker, Ehrlich, and Sjoquist specifications imply theft is a normal activity.[25]

---

[20] For a detailed discussion of wealth equivalence, see Block and Heineke (1973b).

[21] For example, Ehrlich's equations (1.2) and (1.3), p. 525, and Sjoquist's equations (2) and (4), p. 441, all imply that the psychic costs and benefits of both legal and illegal activity are independent of wealth. For a more detailed discussion of this point, see Block and Heineke (1974).

[22] To simplify notation in what follows, $\bar{U}_W$ will be used to represent $d\bar{U}(W^*)/dW^*$.

[23] See fnn. 11 and 7 for a discussion of this point.

[24] For proof of a formally identical proposition, see Sandmo's (1971, pp. 68–69) demonstration of the negative output effects associated with changes in fixed costs.

We now briefly reexamine the other supply effects reported above for the case where C and W are independent.

B. *Payoff Effects* ($C_W=0$)

For the case at hand, equation (5) above becomes

$$(5') \quad \partial T/\partial V = -E\bar{U}_W\theta'/H_{TT} + \theta\partial T/\partial W^0$$

The substitution effect in (5′) is positive and as we have seen under the Arrow-Pratt hypothesis the wealth effect is also positive. Hence, if psychic costs are invariant in wealth and if absolute risk aversion decreases in wealth, then the agent will unambiguously devote more hours to illegal activity as the return to these activities increases.

C. *Enforcement Effects* ($C_W=0$)

Derivative (6) above is of course still

$$(6') \quad \partial T/\partial\delta = -F(\partial T/\partial V)$$

where δ is the mean altering, dispersion preserving, additive shift parameter on the random variable a. But as has been noted, with $C_W=0$ decreasing absolute risk aversion implies $\partial T/\partial V>0$ and therefore increases in the arrest rate will produce an unambiguous deterrent effect on the supply of offenses.

D. *Penalty Effects* ($C_W=0$)

When $C_W=0$, the penalty effect reported in equation (7) becomes

$$(7') \quad \partial T/\partial F = E(\bar{U}_W a)\theta'/H_{TT} + \theta E[\bar{U}_{WW}((V-aF)\theta' - r + C_L - C_T)a]/H_{TT}$$

Since $\bar{U}_W$ and a are each nonnegative random variables, the first term in this expression is negative. In addition, it is easy to show that decreasing absolute risk aversion implies the numerator of the second term is positive.[26] Therefore, both terms are negative and we have the result reported by Becker, p. 177, Ehrlich, p. 529, and Sjoquist, p. 441: Increases in "punishment" unequivocally reduce the incentive to engage in illegal activities. Again the independence of psychic costs and wealth implicit in the Becker, Ehrlich, and Sjoquist models eliminates the ambiguity reported for the general case.

E. *Pure Dispersion Changes*

We now turn our attention to an additional and very interesting parameter shift, a shift discussed by Becker, Ehrlich, and Kolm.

The relation between the offense decision and the degree of certainty with which the penalty is administered has been debated endlessly by criminologists. Well over a century and a half ago, Sir Samuel Romilly, in a series of debates with William Paley, held that not only did certainty of punishment deter criminal activities, but also that certainty of punishment was more crucial then severity. "So evident is the truth of this maxim that if it were possible that punishment could be reduced to an absolute certainty, a very slight penalty would be sufficient to deter almost every species of crime. . . ."[27] We next determine whether the present model contains any implications concerning the deterrent effects of increases in the certainty of punishment.

---

[25] Since the time allocations to income generating activities is fixed, i.e., $N=\bar{N}$, inequality (10) implies $\partial L/\partial W^0<0$.

[26] To see this, let $Z=(V-aF)\theta'-r+C_L-C_T$ and let W_0 be that wealth level such that $Z=0$. We must show $E(\bar{U}_{WW}Za)>0$. If $Z>0$ then $aR<(aR)_0$ where $(aR)_0$ signifies that the product aR is evaluated at W_0 and hence is nonrandom. Therefore, $-Za\bar{U}_{WW}<(aR)_0\bar{U}_WZ$. If $Z<0$ the analogous argument yields the same result. Hence, $-E(\bar{U}_{WW}Za)<(aR)_0E(\bar{U}_WZ)$. But $E(\bar{U}_WZ)$ is the necessary condition for an internal maximum and must be zero. Therefore, $E(\bar{U}_{WW}Za)>0$. Note that if (a) is Bernoulli, risk aversion alone signs (7′). See the Appendix.

A widely utilized method of studying the effects of changes in the dispersion of a random variable consists of using a combination of a multiplicative and an additive parameter shift on the variable in question. The multiplicative shift "spreads" the density, while the additive shift is used to keep the mean of the variable unchanged.[28] To assess the supply effects of a change in the dispersion of punishment, we apply the additive shift parameter to a, say γ, which in turn acts as a multiplicative shift on F. The parameter γ is restricted to ensure $E(aF)$ is unchanged.[29] It is interesting to note that dispersion changes generated in this manner are formally identical to the changes in the probability of arrest "compensated" by changes in the penalty reported by Becker, p. 178, Ehrlich, p. 530, and Kolm, p. 266.[30]

Differentiating the right-hand side of (9) first with respect to T, then with respect to γ and evaluating the result at $\gamma=0$, we have[31]

$$\text{(11)}\quad \partial \mathrm{T}/\partial\gamma = -(F/E(a))\{\mathrm{Cov}\,(\overline{U}_W, a)\theta' + \theta\,\mathrm{Cov}\,[\overline{U}_{WW}((V-aF)\theta' - r + C_{\mathrm{L}} - C_{\mathrm{T}}), a]\}/H_{\mathrm{TT}}$$

Unlike the other comparative static results reported in this section, decreasing absolute risk aversion will not be sufficient to sign (11). For risk-averse agents Cov $(\overline{U}_W, a)$ is positive, but nonlinearities in "ethical costs" $C_{\mathrm{L}}-C_{\mathrm{T}}$, prevent further analysis of the second covariance term. It would seem that this term can be signed only if the function $C(\mathrm{L}, \mathrm{T})$ is linear. An individual for which this condition holds might be said to display *ethical independence*,[32] in which case it can be shown that $\partial R/\partial W<0$ implies the second covariance in (11) is positive and therefore[33]

$$\text{(12)}\qquad \partial T/\partial\gamma > 0$$

Given the preference restrictions which have been enumerated, the model supports the hypothesis that increases in the certainty of punishment will induce disincentive effects. However, this seemingly very plausible result that increases in the certainty of punishment will discourage criminal activity, a hypothesis often accepted as fact, rests upon the assumptions that psychic costs are independent of wealth and that the criminal choice problem is characterized by what Arrow has called "constant stochastic returns." In other words, the deterrent effect of "small" increases in the certainty of punishment is straightforward to establish only when the criminal choice is modeled as a portfolio problem. Thus the results concerning the certainty of punishment reported by Becker, Ehrlich, and Kolm for the case where $f(a)$ is Bernoulli, are not forthcoming in the general case.[34]

---

[27] This debate is reported in Jerome Michael and Herbert Wechsler, p. 250ff.

[28] For example, see Sandmo (1970, 1971), Leland (1972), and Block and Heineke (1973a).

[29] Formally, $dE(a+\gamma)F/d\gamma=0$ and hence $dF/d\gamma = -F/E(a)$ when $\gamma=0$.

[30] This point is also noted in Brown and Reynolds, pp. 512–13. See the Appendix for a more detailed discussion.

[31] Of course, $C_W=0$ in this derivative. If not, there is no possibility of extracting qualitative information.

[32] Formally, if $U=U(\mathrm{L}, \mathrm{T}, W)$, K is a constant and $(\partial W/\partial \mathrm{T})_{U_0}-(\partial W/\partial \mathrm{L})_{U_0}=K$, L, T, $W\geq 0$, the individual is *ethically independent*. Agents whose orderings display ethical independence have the same relative "taste (distaste) for crime" no matter what their wealth may be and no matter how involved they might be in legal and illegal activities. In other words, the agent's ethical considerations are independent of both his wealth and his participation rates in income generating activities. If $K=0$ in the definition, we might say the individual is "ethically neutral." Ethically neutral individuals find legal and illegal activity equally distasteful and in effect combine them under the heading "work." These individuals probably most adequately represent the caricature of "economic man."

[33] Define $Z\equiv(V-aF)\theta'-r+C_{\mathrm{L}}-C_{\mathrm{T}}$. We are to show $\mathrm{Cov}\,(\overline{U}_{WW}Z,a)\equiv E\,[\overline{U}_{WW}Z\,(a-\mu_a)\,]>0$. Note that $Z-\mu_Z\equiv(a-\mu_a)(-F\theta')$ and hence $E[\overline{U}_{WW}Z(a-\mu_a)]\equiv -F\theta' E(\overline{U}_{WW}Z(Z-\mu_Z)]\equiv -F\theta'[E\overline{U}_{WW}Z^2-\mu_Z E\overline{U}_{WW}Z]$. For risk-averse individuals $E\overline{U}_{WW}Z^2<0$ and if $\partial R/\partial W<0$, then $E\overline{U}_{WW}Z>0$ by (10) above. The only remaining unsigned term is μ_Z. If the individual exhibits ethical independence, the necessary condition for a nonzero supply of offenses is $\mu_Z>0$. Therefore, the term in brackets is negative and $\mathrm{Cov}(\overline{U}_{WW}Z, a)>0$.

IV. Summary

We have examined in some detail the individual's choice among two income-generating and time-consuming alternatives, one legal with certain returns and one illegal with stochastic returns. Unlike the existing literature in the area, this problem was formulated in terms of the underlying multiattributed structure of preferences inherent in the decision problem. Utilizing this basic framework and carefully specifying the relationship between the multiattribute problem and its single attribute equivalent, we have shown that the results obtained by previous authors are valid only in special cases. Most significantly, changes in (i) wealth, (ii) the payoff to illegal activity, (iii) enforcement, (iv) punishment, and (v) the degree of certainty surrounding punishment were seen to have no qualitative supply implications under traditional preference restrictions.

Simplifications which may appear to be forthcoming in a "wealth only" model are the result of a failure to fully specify the transformation between the underlying multiattribute model and its single attribute equivalent. Hence, in the area of law enforcement as in taxation, policy recommendations do not follow from theory but rather require empirical determination of relative magnitudes.

APPENDIX

The Bernoulli as Subjective Density

In the analysis above we assumed only the existence of a subjective probability distribution $f(a)$. This is a much more general approach than has been adopted in previous work. Becker's pioneering work in the area and the Ehrlich and Sjoquist extensions assume a is either 1 or 0 with $f(1)=p$ and $f(0)=1-p$; i.e., $f(a)$ is Bernoulli. This implies that the individual makes decisions as if the only possible outcomes are total failure or complete success, although it is difficult to imagine a situation in which the individual would be *either* caught for every offense *or* not caught at all. This is in contrast to the above formulation in which the "arrest rate" may take on any value between 0 and 1 and hence the individual is confronted with a continuum of failure possibilities.[35] He fails on none, on all, or on any fraction of his attempted offenses. While both Ehrlich and Becker seem to suggest that their results are forthcoming for more general densities, as the results above indicate, this is not the case.[36]

To see the implications of this density, define $W' \equiv W^0 + r\mathrm{L} + (V-F)\theta$ and $W'' \equiv W^0 + r\mathrm{L} + V\theta$ and let $f(a)$ be Bernoulli. In this case,

$$\text{(A1)} \quad EU = pU(\mathrm{L}, \mathrm{T}, W') + (1-p)U(\mathrm{L}, \mathrm{T}, W'')$$

which in wealth equivalent form is

$$\text{(A2)} \quad E\bar{U} = p\bar{U}(W' - C(\mathrm{L}, \mathrm{T}, W')) + (1-p)\bar{U}(W'' - C(\mathrm{L}, \mathrm{T}, W''))$$

Note that equation (A2) is the Ehrlich model if $C(\mathrm{L}, \mathrm{T}, W)$ is not subsumed into V and $V-F$.[37]

---

[34] To interpret $\partial\mathrm{T}/\partial\gamma$ in terms of the Becker, Ehrlich, and Kolm results, note that a is increased and F is decreased such that $E(aF)$ is constant. Since $\partial\mathrm{T}/\partial\gamma > 0$, the decrease in F has the greater effect on T. (See the Appendix.)

[35] See Heineke for further discussion of this point.

[36] For example, Ehrlich states, "Although our model has been illustrated for two states of the world, the analysis equally well applies to n states . . ." (p. 528).

[37] There does remain one minor difference between the model in (A2) and the Ehrlich formulation. Ehrlich allows for variable punishment by considering a punishment function $F(\theta)$.

$$\begin{aligned}\text{(A4)}\qquad \partial \mathrm{T}/\partial\gamma = \{&-\overline{U}_W(\overline{W})[(V\theta' - r)(1 - C_W(\mathrm{L}, \mathrm{T}, W')) + C_{\mathrm{L}}(\mathrm{L}, \mathrm{T}, W') - C_{\mathrm{T}}(\mathrm{L}, \mathrm{T}, W')]\\ &+ \overline{U}_W(\underline{W})[(V\theta' - r)(1 - C_W(\mathrm{L}, \mathrm{T}, W'')) + C_{\mathrm{L}}(\mathrm{L}, \mathrm{T}, W'') - C_{\mathrm{T}}(\mathrm{L}, \mathrm{T}, W'')]\\ &- F\theta\overline{U}_W(\overline{W})[C_{\mathrm{L}W}(\mathrm{L}, \mathrm{T}, W') - C_{\mathrm{T}W}(\mathrm{L}, \mathrm{T}, W') - C_{WW}(\mathrm{L}, \mathrm{T}, W')((V - F)\theta' - r)]\\ &- F\theta\overline{U}_{WW}(\overline{W})(1 - C_W(\mathrm{L}, \mathrm{T}, W'))[((V - F)\theta' - r)(1 - C_W(\mathrm{L}, \mathrm{T}, W'))\\ &+ C_{\mathrm{L}}(\mathrm{L}, \mathrm{T}, W') - C_{\mathrm{T}}(\mathrm{L}, \mathrm{T}, W')]\}/G_{\mathrm{TT}}\end{aligned}$$

The essential elements of the preference restrictions underpinning the Becker, Ehrlich, and Sjoquist models may be seen by examining but one of the above supply effects: the effect on illegal activity of a "compensated" increase in the arrest rate. A compensated increase in the arrest rate consists of an increase in the arrest rate compensated by a decrease in the penalty, so that the effect of both changes is to leave the expected punishment pF unchanged. While Becker and Ehrlich employ equal and opposite percentage changes in p and F to accomplish this compensated change, it may also be performed by simply setting

$$d(pF)/dp = F + p(dF/dp) = 0$$

and hence dF/dp is equal to $-F/p$. This latter approach has the advantage of emphasizing the relationship between compensated changes in p and the more general dispersion changes discussed above. Within the Bernoulli framework, the Becker-Ehrlich compensated change is a change in the dispersion of returns to illegal activity.

To proceed, note that

$$\text{(A3)}\quad \partial \mathrm{T}/\partial\gamma = \partial \mathrm{T}/\partial p - (\partial \mathrm{T}/\partial F)(F/p)$$

where $\partial \mathrm{T}/\partial\gamma$ is the effect on illegal activity of a mean preserving (or compensated) change in p, and $\partial \mathrm{T}/\partial p$ and $\partial \mathrm{T}/\partial F$ are the effects on T of changes in p and F, respectively. The individual's optimal level of illegal activity is obtained by maximizing either (A1) or (A2). Since we are interested in isolating the preference restrictions underpinning the Becker, Ehrlich, and Sjoquist results, it is most convenient to pose the decision problem in its wealth equivalent form, (A2). To reduce the notation, define $\overline{W} \equiv W' - C(\mathrm{L}, \mathrm{T}, W')$ and $\underline{W} \equiv W'' - C(\mathrm{L}, \mathrm{T}, W'')$. Equation (A3) may now be written as shown in equation (A4) where $G \equiv p\overline{U}(\overline{W}) + (1-p)\overline{U}(\underline{W})$. We now note that in general, and contrary to the assertions made by both Becker, p. 178, and Ehrlich, p. 530, simple behavior toward risk (*sign* U_{WW}) is not sufficient to establish the qualitative effect of a compensated change in p. That is, the sign of (A4) is not determined by sign U_{WW}—one also needs information on the properties of the "cost" function C. We now show that only in a special case is it possible to infer the sign of (A4) from the sign of $\overline{U}_{WW}$ and also to infer the sign of $\overline{U}_{WW}$ from the sign of (A4).[38]

To see this, consider the special case in which ethical costs are independent of the individual's wealth position, i.e., $C(\mathrm{L}, \mathrm{T}, W) = C(\mathrm{L}, \mathrm{T})$. Under this condition (A4) may be rewritten as follows:

$$\begin{aligned}\text{(A4')}\quad &\partial \mathrm{T}/\partial\gamma\\ &= \{[-\overline{U}_W(\overline{W}) + \overline{U}_W(\underline{W})]\\ &\quad\cdot[(V\theta' - r) + C_{\mathrm{L}}(\mathrm{L}, \mathrm{T}) - C_{\mathrm{T}}(\mathrm{L}, \mathrm{T})]\\ &\quad - F\theta\overline{U}_{WW}(\overline{W})[((V - F)\theta' - r)\\ &\quad + C_{\mathrm{L}}(\mathrm{L}, \mathrm{T}) - C_{\mathrm{T}}(\mathrm{L}, \mathrm{T})]\}/G_{\mathrm{TT}}\end{aligned}$$

Equation (A4′) is the result obtained by Ehrlich and is in fact identical to his expression for a compensated change in p except for the fact that in (A4′) ethical costs have not been aggregated into "net" returns.[39]

It is straightforward to show that the sign of (A4′) is uniquely determined by the sign of $\overline{U}_{WW}$.[40] For example, if the individual is risk averse (A4′) will be positive and a compensated increase in the arrest rate will increase the individual's allocation to illegal activities. In other words, under the condition that *ethical costs are independent of wealth*, a decrease in the dispersion of returns to illegal activities will, *when the density is Bernoulli*, unambiguously lead a risk-averse (risk-preferring) individual to increase (decrease) his supply of such activities. Crucial in this result is the specific density and the independence of ethical costs and wealth. As we have shown above, if the density is not Bernoulli and/or ethical costs are not independent of wealth, simple behavior toward risk is not sufficient to establish the effect of mean preserving dispersion changes.

---

[38] More precisely in the Becker, Ehrlich, and Sjoquist models, $\partial T/\partial\gamma \gtreqless 0$ iff $U_{WW} \lesseqgtr 0$.

[39] See Ehrlich, fn. 13, p. 530.

[40] To see this, note that $((V-F)\theta' - r + C_{\mathrm{L}}(L, \mathrm{T}) - C_{\mathrm{T}}(\mathrm{L}, \mathrm{T})) < 0$ and $(V\theta' - r + C_{\mathrm{L}}(\mathrm{L}, \mathrm{T}) - C_{\mathrm{T}}(\mathrm{L}, \mathrm{T})) > 0$ by the first-order condition and $G_{\mathrm{TT}} < 0$ by the second-order condition. Therefore the sign of the first term on the right-hand side of (A-4′) will be determined by the sign of $\overline{U}_{WW}$. Since the sign of this term will be opposite that of $\overline{U}_{WW}$, the sign of $\overline{U}_{WW}$ uniquely determines the sign of (A-4′). In fact, with $C_W = 0$, $\partial \mathrm{T}/\partial\gamma \gtreqless 0$ iff $\overline{U}_{WW} \lesseqgtr 0$.

REFERENCES

M. G. Allingham and A. Sandmo, "Income Tax Evasion: A Theoretical Analysis," *J. Publ. Econ.*, Nov. 1972, *1*, 323–38.

G. Becker, "Crime and Punishment: An Economic Approach," *J. Polit. Econ.*, Mar./Apr. 1968, *78*, 169–217.

M. K. Block, "An Economic Analysis of Theft with Special Emphasis on Household Decisions Under Uncertainty," unpublished doctoral dissertation, Stanford Univ. 1972.

——— and J. M. Heineke, "A Comment on Uncertainty and Household Decisions," *Rev. Econ. Stud.*, Oct. 1972, *39*, 523–25.

——— and ———, "The Allocation of Effort Under Uncertainty: The Case of Risk Averse Behavior," *J. Polit. Econ.*, Mar./Apr. 1973, *81*, 376–85.

——— and ———, "Multiattributed Preferences and Wealth Equivalents: The Case of Illegitimate Activity," working pap. no. 16, Grad. Sch. Bus., Univ. Santa Clara 1974.

——— and ———, "The Criminal Choice and Trading Ethics," paper presented at the European Meetings of the Econometric Society, Oslo, Norway, 1973.

——— and R. Lind, "A Choice Theoretic Analysis for Crimes Punishable by Imprisonment," *J. Legal Stud.*, forthcoming.

W. Brown and M. Reynolds, "Crime and Punishment: Risk Implications," *J. Econ. Theory*, Oct. 1973, *6*, 508–14.

I. Ehrlich, "Participation in Illegitimate Activities: A Theoretical and Empirical Investigation," *J. Polit. Econ.*, May/June 1973, *81*, 521–67.

J. M. Heineke, "A Note on Modeling the Criminal Choice Problem," *J. Econ. Theory*, forthcoming.

R. L. Keeney, "Risk Independence and Multiattributed Utility Functions," *Econometrica*, Jan. 1973, *41*, 27–34.

S. Ch. Kolm, "A Note on Optimum Tax Evasion," *J. Publ. Econ.*, July 1973, *2*, 265–70.

H. Leland, "Savings and Uncertainty: The Precautionary Demand for Saving," *Quart. J. Econ.*, Aug. 1968, *82*, 465–73.

———, "Theory of the Firm Facing Uncertain Demand," *Amer. Econ. Rev.*, June 1972, *62*, 278–91.

J. Michael and H. Wechsler, *Criminal Law and its Administration*, Chicago 1940.

J. Mossin, "Taxation and Risk Taking: An Expected Utility Approach," *Economica*, Feb. 1968, *35*, 74–82.

———, "Aspects of Rational Insurance Purchasing," *J. Polit. Econ.*, July/Aug. 1968, *76*, 553–68.

M. O. Reynolds, "Crimes for Profit: The Economics of Theft," unpublished doctoral dissertation, Univ. Wisconsin 1971.

A. Sandmo, "The Effect of Uncertainty on Savings Decisions," *Rev. Econ. Stud.*, July 1970, *37*, 353–60.

———, "On the Theory of the Competitive Firm Under Price Uncertainty," *Amer. Econ. Rev.*, Mar. 1971, *61*, 65–73.

C. Schrag, "Crime and Justice: American Style," Crime and Delinquency Issues, pub. no. HSM-72-9052, Nat. Inst. Mental Health, Washington 1971.

B. Singh, "Making Honesty the Best Policy," *J. Publ. Econ.*, July 1973, *3*, 257–63.

D. L. Sjoquist, "Property Crime and Economic Behavior: Some Empirical Results," *Amer. Econ. Rev.*, June 1973, *63*, 439–46.

G. Stigler, "The Optimum Enforcement of Laws," *J. Polit. Econ.*, Mar./Apr. 1970, *78*, 526–36.

Monopoly Pricing and Capacity Choice Under Uncertainty

By ROBERT A. MEYER*

The analysis of pricing and investment problems for a monopoly under certainty has received considerable attention in the guise of the peak-load pricing model with notable contributions by Marcel Boiteux, Herbert Mohring, Peter Steiner, and Oliver Williamson. Inquiry into the impact of rate of return regulation under certainty has proceeded along an independent track, beginning with the well-known paper by Harvey Averch and Leland Johnson, until the peak-load and regulation issues were merged by Elizabeth Bailey.

In contrast to these analyses of monopoly behavior under certainty, models of decision making under uncertainty have generally dealt with only a single product firm and an unregulated setting. While one might argue at some length over the appropriate definition and measurement of risk, its presence obviously needs to be recognized in the process of deriving pricing and investment policies for both regulated and unregulated firms.

This paper extends the monopoly pricing and investment problem under uncertainty to the case of multiple outputs[1] and at the same time uses an alternative method for representing and analyzing the effects of risk. Since one may regard a multiple output setting as applying to the same product at different points in time, the model presented here is also a generalization of the peak-load problem to allow for stochastic demands. Another and perhaps more interesting interpretation is to consider multiple outputs as sales to different groups of customers at a single point in time. This latter interpretation will be shown to lead to a new way of viewing pricing under risk which leads to a rationale for price discrimination quite different from the traditional one under certainty. In particular, this leads to implications for differential pricing between "steady customers" vis-à-vis other customers.

Differing risk preferences of the decision maker may obviously alter the implied pricing and investment rules, but the generality and elegance of using an expected utility criterion sometimes prohibits one from obtaining clear comparisons. It will be shown that use of a chance-constrained programming formulation leads to several simplifications, both in tractability and ease of practical application.

The following discussion is developed in four parts. Section I describes the basic model to be used in the remainder of the discussion and summarizes the implied pricing and investment rules under certainty. Section II extends the discussion to incorporate uncertain demands and discusses the structure of alternative monopoly models under risk. Results are developed for the risk-neutral case using both an expected profit criterion and the expected "welfare" criterion. Section III introduces the use of chance-constrained programming as a vehicle for representing non-linear risk preferences. Section IV suggests an alternative view of pricing under risk relying on an analogy between portfolio selection and choosing prices for multiple groups of customers. A conclusion summarizes the main points.

---

* Associate professor of applied economics, University of California, Berkeley. I would like to express my appreciation to M. Bruce Johnson, John A. Carlson, Morton Kamien, and anonymous referees for helpful comments. Research support by the Institute of Public Utilities, Michigan State University, is gratefully acknowledged.

[1] Phoebus Dhrymes also considers a multiple output case, but in the Dhrymes model, prices are random and a quadratic expected utility criterion is employed.

I. Riskless Monopoly Pricing

The basic model of a monopoly firm under certainty selling to n groups of customers (or the peak-load problem with time divided into n periods) has been developed using the following set of assumptions:

A1. A demand cycle (such as a day, month, or year) is divided into n periods with a known deterministic demand, $D_i(p_i)$, $i=1, \dots, n$ which depends only on the price charged in period i; $dD_i(p_i)/dp_i<0$.

A2. The admissible set of pricing rules consists of those prices which are constants over inframarginal units *and* over individuals in the same period.

A3. The "investment problem" is simply to choose once and for all "the" optimal capacity, which is regarded as completely divisible.

A4. Production costs are composed of two separable parts: a variable cost per unit of *output per demand cycle* b, and a capacity cost of β per unit of *capacity per demand cycle* (i.e., per period).[2]

A5. The firm's objective is to choose capacity level and prices such as to maximize the sum of consumers' surplus, net revenue to the firm, and producers' surplus.[3]

Optimal pricing and investment rules for a welfare maximizing objective are obtained from

$$(1)\quad \underset{p_1, p_2, \dots, p_n, Q}{\text{Maximize}}\; W = \sum_{i=1}^{n} (p_i - b) D_i(p_i) - \beta Q + \sum_{i=1}^{n} \left\{ \int_0^{D_i(p_i)} [D_i^{-1}(x) - p_i] dx \right\}$$

subject to

$$(2)\qquad D_i(p_i) \le Q \qquad i = 1, \dots, n$$

$$(3)\qquad Q \ge 0; \quad p_i \ge 0, i = 1, \dots, n$$

with the corresponding model for a profit maximizer being the same except for the omission of the third term in the objective. Table 1 provides a comparison of the pricing and investment rules for both types of objectives.[4] The welfare maximizer always charges a peak price higher than its off-peak price, while this is not necessarily true for a profit maximizer. A

---

[2] For the deterministic setting, a static model can be linked to a dynamic model of the firm and thus provide a more explicit interpretation of the capacity cost β. Modern capital theory provides a direct link for the reversible capital model—that is, a model in which the firm may purchase or sell capital at a price c_t per unit of capacity at time t. If we let δ_t denote the amount of the capital stock at the beginning of period t, K_t that is available at the end of the period, and r_t the one-period interest rate for period t, then it has been shown by Kenneth Arrow and Dale Jorgenson that the user cost of capital in period t is $\beta_t = c_t(1+r_t) - \delta_t c_{t+1}$ per unit (ignoring taxes). The foregoing remarks are not intended to justify a reversible capital model for a public utility, since such a model may be particularly undesirable as a description of the institutional environment and capital markets in which a public utility must deal; however, this indicates part of the nature of a dynamic model of the firm for which it is formally legitimate to solve the dynamic problem as a sequence of static problems. Treating average operating and capacity costs as constants will make it possible to bring the stochastic elements into much sharper focus. In addition, assumption A4, together with the fact that demand is deterministic and the problem setting is static, removes any economic rationale for excess capacity; thus $Q = \max_i \{D_i(p_i)\}$. This point is worth noting since the introduction of stochastic demand and/or the introduction of reliability of service constraints provide an economic motivation for having optimal capacity greater than expected demand.

[3] Since production costs are constant over inframarginal units, producers' surplus is zero.

[4] In the deterministic setting, results simplify quite readily since one knows the optimal price, p_i^*, for peak periods is such that $p_i^* = D_i^{-1}(Q^*)$ where Q^* is optimal capacity.

TABLE 1—COMPARISONS OF PRICING AND INVESTMENT RULES[a]

Criterion	Peak Price	Off-Peak Price	Investment Rule
Profit Maximization[b]	$p_i = b - D_i/D_i' + v_i$	$p_i = b - D_i/D_i'$	$\sum_{i=1}^{n} v_i = \beta$
	or	or	
	$p_i = \frac{\eta_i}{(1+\eta_i)}(b + v_i)$	$p_i = \frac{\eta_i}{(1+\eta_i)} b$	
Welfare Maximization	$p_i = b + v_i'$	$p_i = b$	$\sum_{i=1}^{n} v_i' = \beta$

[a] For the profit maximizing objective the v_i are the Lagrangian multipliers for the excess demand constraints $Q - D_i(p_i) \geq 0$. The symbols v_i' denote the corresponding terms for the welfare maximizing objective. For either objective, v represents the quasi rent (or marginal revenue product of capacity) from serving the ith group or period. The symbol η_i denotes the familiar price elasticity of demand, $\eta_i = dD_i/dp_i \cdot p_i/D_i$. The term peak is used for the case where the multiplier is strictly positive; however, a zero multiplier does *not* imply "off-peak" quantity demanded is less than "peak" quantity demanded since a multiplier may be zero even though the constraint holds with equality.

[b] Peak prices are not necessarily higher than off-peak prices.

profit maximizer's off-peak is equal to or greater than the welfare maximizer's off-peak price, but the profit maximizer's peak price may be *less* than the welfare maximizer's peak price. No general statement can be made concerning a profit maximizer's optimal capacity vis-à-vis a welfare maximizer's unless one knows a profit maximizer's peak prices are equal to or greater than the welfare maximizer's *in every period* (or for all customers).

II. Monopoly Pricing Under Risk

This section develops the stochastic choice problem based on both expected profit maximization and expected welfare maximization criteria. Since risk neutrality implies the firm is indifferent to risk, a more flexible scheme for treating risk-averse settings is presented in the subsequent two sections.[5] The introduction of risk—in particular, random components in demand—requires a closer look at the structure of the decision problem. What is the objective to be? What are the decision variables? How does randomness enter into the demand relationship?

---

[5] Aside from introducing risk per se, random demand poses another problem: How is demand to be rationed if realized demand exceeds capacity? This question is directly tied to the choice of decision variables. Since many of the actual settings in which the peak-load question arises are found in firms selling air transport service, electricity, telephone, natural gas, etc., it seems most plausible to treat price and capacity as decision variables. Even for other types of firms selling to multiple groups of customers, price setting seems to be a pervasive business practice. By requiring that price and capacity be chosen *prior* to demand realization there is a possibility that actual demand at the chosen price may exceed capacity, thus raising the rationing issue. Michael Visscher in commenting on an earlier contribution by Gardner Brown and M. Bruce Johnson notes that the rationing rule adopted will affect the implied optimal prices and output. This is true, however, only when one uses a welfare criterion. Expected profit maximization or even the choice of a utility function defined on profits will not lead to such an ambiguity since only the amount the firm receives "counts" and this is received from consumers who are at least willing to pay the previously named price. How much more some consumers that were denied might have been willing to pay is not relevant. Since Visscher's random rationing scheme will not alter inferences drawn from the expected profit maximization criterion one may assume this is followed; however, to preserve comparability with the deterministic results of Steiner, Williamson, and others, and with the results of Brown and Johnson, it will be assumed output is allocated to those having the highest marginal valuations first when the welfare criterion is employed.

Both the expected profit maximization and expected welfare maximization settings are developed on the basis of the following assumptions:

B1. For each time period i demand is given by $D_i(p_i)=\bar{D}_i(p_i)+u_i$ where $\bar{D}_i$ is the deterministic (or riskless) part with $-\infty < d\bar{D}_i/dp_i < 0$, and u_i is a random part defined to take values so that $0 \leq D_i < \infty$. It is assumed that u_i is drawn from a covariance stationary stochastic process with finite variance. No specific assumption is made as to the mean of u_i.

B2. $\partial D_j/\partial p_i = 0$ for all $j \neq i$ (thus, all cross-price elasticities are zero). This is certainly in conformity with reality if one thinks of demands as corresponding to different classes of customers, for example, residential vs. commercial.

B3. Total cost is composed of two parts; a variable cost of b per unit of output and a capacity cost of β per unit of capacity. The firm may adjust output *after* demand is known, but only up to the limit imposed by capacity which must be chosen ahead of time. Long-run investment choice thus enters as an upper limit on short-run production flexibility.

B4. The random component of demand possesses a density function $g_i[\cdot]$ and a distribution function $G_i[\cdot]$ independent of prices.

B5. The quantity demanded at a given price depends on the price per unit and does not depend on the "quality" of service[6] (for example, the reliability of service).

Results at each step are contrasted with the corresponding results for deterministic setting where possible.

The objective for the expected profit maximizer[7] is

$$(4)\quad E[\pi] = \sum_{i=1}^{n} \int_{-\bar{D}_i(p_i)}^{Q-\bar{D}_i(p_i)} g_i(u_i)(p_i - b) \cdot [\bar{D}_i(p_i) + u_i]du_i + \sum_{i=1}^{n} (p_i - b) \cdot Q\{1 - G_i[Q - \bar{D}_i(p_i)]\} - \beta Q$$

The first term in (4) is expected net operating revenue when actual demand is equal to or greater than zero and equal to or less than capacity, and the second term is expected net operating revenue if actual demand exceeds capacity. The last term is capacity cost. A positive optimal level of investment requires

$$(5)\quad \sum_{i=1}^{n} \{1 - G_i[Q - \bar{D}_i(p_i)]\}(p_i - b) = \beta$$

and the corresponding pricing rule for

Since the rationing question has drawn attention to the welfare criterion, it might be worthwhile to emphasize two other inherent drawbacks. Attempting to apply such a criterion in practice requires obtaining an estimate of *marginal* demand schedules, i.e., schedules of marginal willingness to pay. Even if the correct marginal schedule is employed in the theoretical part of the analysis, the prospects for empirically determining it seem somewhat less than sparkling. A second, though related point, is that the objective function typically employed to represent "welfare" adds the firm's net revenue and the uncaptured portion of consumers' valuation of output with *equal* weights. While there is probably no such thing as a "correct" weighting scheme, still it does seem unlikely that firms would be inclined to count captured and uncaptured potential revenues on an equal footing. At the very least one can view the prospect with some degree of skepticism.

[6] For an alternative certainty model which explicitly introduces "quality of product" into a regulated framework, see James Rosse.

[7] Even stated as an unconstrained maximization problem it is by no means clear that such a formulation is a well-formed maximum problem. If the objective function is strictly concave in p_i and Q, then a unique solution exists (although these values are not necessarily guaranteed to be nonnegative). Following the literature, it is assumed that the objective function here and in subsequent formulation is strictly concave in its variables and attains its maximum for nonnegative values of the variables. Generally, such an assumption will be found to implicitly restrict other functions in the problem.

positive optimal prices is implicitly defined by

$$(6)\quad \left\{(p_i - b)\frac{d\overline{D}_i}{dp_i} + \overline{D}_i\right\} G_i[Q - \overline{D}_i(p_i)] + \int_{-\overline{D}_i(p_i)}^{Q-\overline{D}_i(p_i)} u_i g_i(u_i)du_i + Q\{1 - G_i[Q - \overline{D}_i(p_i)]\} = 0 \qquad i = 1, \ldots, n$$

Equation (5) has a natural interpretation which is a direct extension of the certainty case. Under uncertainty one does not know *in advance* which period (or group of customers) will represent a peak demand. The left-hand side of (5) represents the demand for capacity at the margin and $p_i - b$ represents the quasi rent contributed by the ith period consumers. These quasi rents are not summed directly as in the certainty case, but instead are weighted by the probability that period i is a peak period, i.e., by $1-G_i[Q-\overline{D}_i(p_i)]$ where $G_i[\cdot]$ is the probability that demand in the ith period does not exceed capacity. When demand is stochastic the sharp distinction between peak and off-peak periods disappears whereas it was a center of attention in a riskless setting.

Fortunately explicit evaluation of (6) in order to compare pricing rules is not necessary since the pricing rules for a welfare maximizer take a much simpler form. Comparisons will be delayed until the results for the welfare maximizer are developed below.

The objective function for an expected welfare maximizing firm is shown in equation (7). The first term represents net operating revenue for the firm when realized demand does not exceed capacity, the second is net operating revenue when realized demand exceeds capacity, the third is capacity cost, and the fourth and fifth terms represent the consumers' surpluses corresponding to the first two terms.

From (7) the necessary condition for a positive optimal level of capacity is

$$(8)\quad \sum_{i=1}^{n} \{1 - G_i[Q - \overline{D}_i(p_i)]\}(p_i - b) - \beta + \sum_{i=1}^{n} \int_{Q-D_i(p_i)}^{\infty} g_i(u_i) \cdot [\overline{D_i}^{-1}(Q - u_i) - p_i]du_i = 0$$

Notice that this reduces to the Brown and Johnson result for $n=1$. The third term in (8) is equal to or greater than zero for all Q and p_i. The first and last terms of (8) comprise the demand schedule for capacity for a welfare maximizing firm (at the margin) whereas only the first is used for a profit maximization firm.

Optimal prices for a welfare maximizer in a stochastic setting are found from the implicit pricing rules which satisfy

$$(9)\quad (p_i - b)\frac{d\overline{D}_i}{dp_i}\{G_i[Q - \overline{D}_i(p_i)] - G_i[-\overline{D}_i(p_i)]\} = 0 \quad i = 1, \ldots, n$$

Provided $d\overline{D}_i/dp_i$ and the third term are

$$(7)\quad E[W] = \sum_{i=1}^{n}\left\{\int_{-\overline{D}_i(p_i)}^{Q-\overline{D}_i(p_i)} g_i(u_i)(p_i-b)[\overline{D}_i(p_i)+u_i]du_i + (p_i-b)Q\{1-G_i[Q-\overline{D}_i(p_i)]\}\right\} - \beta Q + \sum_{i=1}^{n}\left\{\int_{-\overline{D}_i(p_i)}^{Q-\overline{D}_i(p_i)} g_i(u_i)\int_{p_i}^{\overline{D}_i^{-1}(-u_i)} [\overline{D}_i(x) + u_i]dx du_i + \int_{Q-\overline{D}_i(p_i)}^{\infty} g_i(u_i)\left\{\int_{p_i}^{\overline{D}_i^{-1}(-u_i)} [\overline{D}_i(x)-u_i]dx - \int_{p_i}^{\overline{D}_i^{-1}(Q-u_i)} [\overline{D}_i(x)+u_i-Q]dx\right\} du_i\right\}$$

not zero, the optimal price for the welfare maximizer is the same in each period and is equal to the marginal operating cost b. A welfare maximizer's optimal prices are equal to or less than the corresponding ones for a riskless setting. In addition, the expected profit maximizer's prices are at least as high as (or higher than) the expected welfare maximizer's in every period (or to every group of customers)—a result which is not necessarily true in a riskless setting.

Since the optimal stochastic price of b holds for all periods, one obtains the Brown-Johnson result for $n=1$ as a special case. Brown and Johnson's conclusion that the firm's net revenue will be negative regardless of which (positive) level of capacity is chosen thus carries over to multiple demand groups also. Inducing a firm to follow a behavior rule which leads to such an implausible result may be yet another argument for ignoring such a "welfare" choice criterion, but there is an element omitted from the Brown and Johnson argument which may rectify the pricing aspect of the problem. While it is clear that an expected profit maximization criterion ignores risk aspects, the use of an expected welfare criterion does also. This additional part of the problem is developed in the next section.

III. Pricing With Risk Limitations

The earlier formulation of the expected welfare maximization model by Brown and Johnson for a single period led to two major conclusions: 1) optimal price equals marginal (and average variable) cost, thus a firm's net profits will always be negative; and 2) stochastic optimal capacity exceeds capacity in the corresponding deterministic setting. While the first conclusion is reached under fairly general conditions employing an additive random error term, the second conclusion was developed only for the case of a linear demand function. For all of their results Brown and Johnson employ a risk neutral criterion, thus the burden of any inherent risk falls on consumers and/or stockholders. In commenting on the Brown and Johnson formulation, Ralph Turvey (1970) emphasized that in a practical situation price must necessarily be specified prior to knowing what actual demand will be and that such prices must be chosen sufficiently high to meet specified standards of system reliability.

These observations point to a more practical way of structuring the problem by introducing chance-constrained programming. Aside from its practical aspects, a chance-constrained programming formulation allows one to directly introduce a "reliability of service" or risk constraint and analyze the effects of certain types of risk-averse behavior without the explicit introduction of a utility function. In addition, it provides a formal basis for analyzing the issue raised by Brown and Johnson (1970) when they state: "To price so high that there is rarely (never?) excess demand implies persistent excess supply (idle capacity) and an *ad hoc* decision by the public utility to remove the risk of failure. We might have more confidence in such a policy if it was derived from an optimizing model" (p. 490). It will be shown that such a price and investment policy by a firm may be far from *ad hoc*—in fact, it is optimal since the degree of failure risk is not necessarily a parameter subject to choice by the firm. Reliability of service or quality standards are frequently specified exogenously by regulatory agencies.

A chance-constrained formulation admits a clearly tractable solution which requires very few specific assumptions in order to determine the qualitative effects of a stochastic setting. An original problem is appended with a set of probabilistic constraints of the form

$$(10)\quad Prob\,\{D_i(p_i) > Q\} = 1 - G_i[Q - \overline{D}_i(p_i)] \leq \epsilon_i > 0, \quad i = 1, \ldots, n$$

where Q is capacity.[8] Constraints of this form would appear to be the natural extension to a stochastic setting of the constraints used in a riskless setting (see equation (2)), but prior work on choice under risk has not included this element.

A firm may price *as if* demand is not random by using an augmented demand schedule which satisfies (10). This is perhaps more easily indicated in a diagram. A new demand schedule $\tilde{D}_i$ is constructed in such a manner that, for each price, constraint (10) is satisfied (Figure 1). To gain some impression of the qualitative effect of ignoring such constraints, suppose Q^0 is the deterministic optimal capacity and one changes *only* prices in order to meet the risk constraints. Periods in which demand is relatively inelastic will require fairly substantial increases in price while periods with more elastic demands will require smaller price increases. In either case, it is clear that prices in each period must be raised above those that would prevail in the deterministic case, p_1^0 and p_2^0. In general, prices found in this manner will not, however, be the stochastic optimal prices, but such prices are the *minimum* prices consistent with the risk constraints for any given level of capacity. If the probability densities are symmetric about their mean, then the optimal deterministic prices p_1^0 and p_2^0 would imply a 50 percent risk of "system failure" in each period!

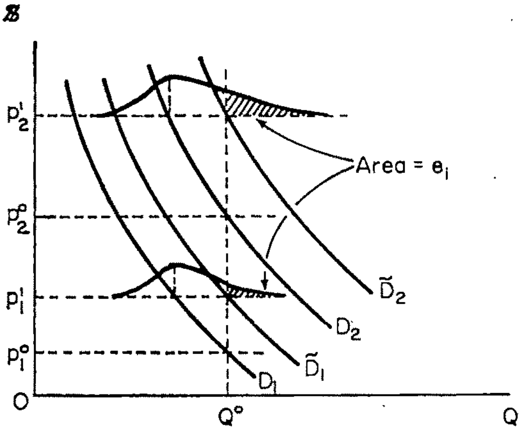


FIGURE 1

The smaller one sets ϵ_i (i.e., the smaller the risk of excess demand permitted), the larger will be the associated price increases. Solving for a price policy subject to (10) meets the criticism raised by Turvey (1968) against omission of a stochastic element: ". . . the price in any period should be such that the mean expectation of demand is *less* than available capacity. . . . This dimension of the optimal price and capacity problem has been ignored by most authors" (p. 105). The variance of demand at any price may, of course, also be dependent upon price and may be different between periods without changing the character of the solution for a given level of capacity. In a similar manner, nothing basic is altered by allowing asymmetric probability distributions.

The most straightforward analysis of the effect of imposing risk or quality constraints is probably to use the optimal solution for prices and capacity in the unconstrained stochastic setting as a refer-

---

[8] Implicitly the constraints in this form assume all demand covariances are zero and $G_i[0]=0$. The probabilistic contraints can also be written in an equivalent riskless form:

$$Prob\,\{D_i(p_i) > Q\} \leq \epsilon_i$$

is equivalent to

$$Prob\left\{\frac{D_i(p_i) - \overline{D}_i(p_i)}{\sigma_i} > \frac{Q - \overline{D}_i(p_i)}{\sigma_i}\right\} \leq \epsilon_i$$

which is satisfied by

$$Q \geq \overline{D}_i(p_i) + N_i\sigma_i$$

where N_i is the number of standard deviations above the mean, $\overline{D}_i(p_i)$, needed to reduce the area in the upper tail of the distribution to ϵ_i.

ence point. Since the objective function cannot achieve a higher value when the constraints are added, a firm seeks to meet the constraints with the smallest (possibly zero) reduction in the objective function value. The firm has at its disposal two instruments—change prices or change capacity—and both instruments have costs attached. One could meet a reliability constraint by increasing capacity alone, which would cost β per unit, or by raising prices, which would "cost" the marginal revenue foregone. The economically optimal response will be to change prices alone provided there is a set of price changes which have a total marginal revenue loss of less than β; otherwise, some combination of price changes and a change in capacity will be necessary to meet the constraint.

Maximization of expected profits or expected welfare subject to the risk constraints does not imply risk neutrality. Introduction of the risk limitations simply implies firms are risk neutral over the sets of prices and output which meet or stay within the risk limits. Reducing the ϵ_i is essentially equivalent to decreasing the degree of excess demand risk embodied in the resulting price and output choices. Although this is a less general approach than using the expected utility criterion employed by Hayne Leland, it is more tractable.

Basic results using the chance-constrained formulation are obtained by appending constraints (10) to (7) for the expected welfare criterion or to (7) after omitting the last two terms for the expected profit criterion. The investment rule (i.e., necessary condition) for an expected profit maximizer is

$$\text{(11)} \quad \sum_{i=1}^{n} \{1 - G_i[Q - \bar{D}_i(p_i)]\}(p_i - b) - \beta - \sum_{i=1}^{n} \lambda_i g_i[Q - \bar{D}_i(p_i)] = 0$$

Assuming the feasible set defined by (10) is convex, the multipliers associated with each constraint, λ_i, are nonpositive and thus the last term in (11) is nonnegative. Optimal constrained capacity is thus equal to or greater than optimal unconstrained capacity for an expected profit maximizing firm when compared at the *same* set of prices.

A comparison between price and capacity in the constrained versus the unconstrained setting can be sketched for the case $n=1$ when G is concave and strictly increasing and $\bar{D}$ is convex, since G is then concave in Q and p and the set $\{(Q, p); G[Q-\bar{D}(p)] \geq 1-\epsilon\}$ is convex. The combinations of Q and p along the lower boundary of this set form an iso-risk contour. Figure 2 indicates the nature of the feasible set, the iso-expected profit coutours, and an example of a possible solution where the subscripts refer to constrained (C) and unconstrained (U) stochastic solutions.

The negative of the multipliers on the risk constraints measure the implicit economic price (per unit) of risk reduction —essentially the maximum insurance premium for risk removal. As ϵ_i decreases, constrained expected profits decrease over their risk-neutral level which simply re-

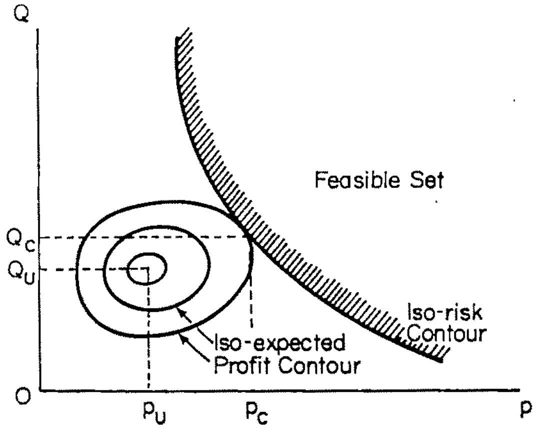


FIGURE 2

flects the fact that the firm would have been paid for assuming risk. The real cost of the risk undertaken "by the firm," however, might well be borne by customers who cannot satisfy their demand while it is the firm who stands to reap any potential benefits (apart from goodwill loss). Obviously one needs a more elaborate problem structure to develop the concept of optimal sharing of risks between the firm, consumers, and perhaps shareholders as well.

For the expected welfare criterion[9] subject to the risk constraints, one obtains the investment rule

$$(12)\quad \sum_{i=1}^{n}\{1-G_i[Q-\overline{D}_i(p_i)]\}(p_i-b) + \sum_{i=1}^{n}\left\{\int_{Q-D_i(p_i)}^{\infty} g_i(u_i) \cdot[\overline{D}_i^{-1}(Q-u_i)-p_i]du_i\right\} - \beta - \sum_{i=1}^{n}\lambda_i g_i[Q-\overline{D}_i(p_i)] = 0$$

Comparison of the optimal prices under constraint with those which would prevail without constraint indicates the prices for the expected welfare maximizer must be equal to or greater than those when no constraints are imposed. This implication follows from the fact that the unconstrained prices were all equal to b and it is never optimal to sell at less than minimum average variable cost. The optimal pricing rule is given by

$$(13)\quad p_i = b - \lambda_i g_i[Q-\overline{D}_i(p_i)] / \{G_i[Q-\overline{D}_i(p_i)] - G_i[-\overline{D}_i(p_i)]\}, \quad i = 1, \ldots, n$$

The term in braces represents the probability that actual demand is equal to or greater than zero but does not exceed capacity. With at least one constraint binding, a higher price using risk constraints implies capacity could be reduced over what would have been required to meet the constraint while charging a price of b.

IV. Risk-Adjusted Price Differentials

The discussion of both the riskless setting and pricing in a risky setting, but without risk constraints, focuses the main attention on pricing and investment rules which are "efficient" in the traditional economic sense. Once one has added risk constraints as in Section IV, a second type of efficiency suggests itself: risk efficiency.

The variance of aggregate demand depends on the prices charged each set of customers since these will alter their contribution to total demand. In addition, the variance of demand may change with price for any class of customers or there may be covariances between the demands of various customers. From basic probability rules the variance of total demand is the sum of the variances of each individual demand plus twice the sum of each pair of covariances. The possibility this presents is analogous to a portfolio selection problem. The variance of aggregate demand, and thus total output as well as profit or revenue risk, may be altered by the *mix* of prices. This suggests a rationale for price discrimination which is entirely different from the traditional one (see Kahn, p. 142).

Once one focuses on the effect on profit or total output variability associated with pricing choices one can find an immediate reason for the apparent dependence of optimal pricing rules on the way in which randomness is introduced. In an earlier exploration of the effect of demand uncertainty on monopoly pricing in riskless versus risky settings, Samuel Karlin and Charles Carr noted that: "The optimal price under multiplicative uncertainty is always greater, and under additive uncertainty always less, than the optimal price set in the equivalent deterministic monopoly models" (p. 160). In both their dynamic and one-period formulations, demand was the only random element. No explanation for this apparent sensitivity of the conclusions to model specification is offered by either Karlin and Carr or in subsequent analysis.

---

[9] It was suggested to me by Morton Kamien and also noted by Edwin Mills that a "correct" objective which is purported to measure "welfare" should include measures of product quality in the objective function. When included in the objective, "quality" could be under control of the firm as in Rosse.

When one adds the risk-efficiency notion, a partial explanation for these results emerges.[10] The variance of total revenue in the additive error model is[11]

$$(14)\qquad Var\,[TR] = \sigma^2 = \sum_{i=1}^{n} p_i^2\, Var\,[D_i(p_i)] = \sum_{i=1}^{n} p_i^2\sigma_i^2$$

where $D_i(p_i)=\bar{D}_i(p_i)+u_i$, $E[u_i]=0$, $E[u_i^2]=\sigma_i^2$. Since prices are the objects of choice one would obviously like to choose these so that (14) is a minimum for a given level of expected profits. The effect of altering the ith price is

$$(15)\qquad \frac{\partial\sigma^2}{\sigma p_i} = 2p_i\sigma_i^2 > 0$$

Thus it is easy to see that total risk decreases as one decreases prices in the additive case. When one moves from a riskless setting to a risky one, risk is lowered by lowering prices over the corresponding riskless ones.

In contrast, the multiplicative error model (i.e., $D_i(p_i)=u_i\bar{D}_i(p_i)$, $E[u_i]=1$, $Var[u_i]=\sigma_i^2$) also has an associated total revenue (or profit) risk element given by

$$(16)\qquad Var\,[TR] = \sigma^2 = \sum_{i=1}^{n} p_i^2\bar{D}_i^2(p_i)\sigma_i^2$$

Changing prices here also affects total profit risk.

$$(17)\qquad \frac{\partial\sigma^2}{\partial p_i} = 2p_i\bar{D}_i^2(p_i)[1+\eta_i]\sigma_i^2$$

where $\eta_i=p_i/\bar{D}_i\cdot d\bar{D}_i/dp_i$. Since $\eta_i<0$ the term in brackets is generally negative, thus risk is decreased by increasing prices when one moves from a riskless world to one characterized by multiplicative demand uncertainty.

Risk-efficient pricing can be characterized quite clearly in terms of the familiar mean-variance portfolio model. For every level of expected profits one defines the associated risk-efficient price set as the set of prices such that the variance of total profit is a minimum. The need for such an additional efficiency concept seems rather clear since a firm faced with selling multiple products (or to multiple classes of customers) could raise the same level of total revenue from a wide array of pricing choices along its iso-revenue surfaces, but even though total revenue remained constant, total risk would not.

When the σ_i^2 are different, then (even though they do not depend on price) the choice of different price sets changes the proportions of total demand (i.e., the proportions of each asset in the portfolio); thus total risk changes. For the perfectly competitive firm even though there exist iso-expected profit surfaces reflecting price/risk choices, the firm cannot avail itself of such tradeoffs since it acts as a price taker. Such tradeoffs are available to a monopolist.

The risk-constrained structure of the problem might be reformulated in terms of total profit risk rather than excess demand risk which brings out a risk premium facet to an optimal pricing structure. Assume a firm wishes to maximize expected profits, but with an upper limit $\bar{\sigma}^2$ on the variance of total profits:

---

[10] Since Karlin and Carr use only an expected profit criterion, their results do not arise from the risk aspect of pricing emphasized here.

[11] This and the following expressions for total revenue variance assume demand covariances are zero. In addition, the variance in demand is used in each expression which implies production is adjusted within a given period to meet realized demand. If production and/or capacity must be chosen ahead of time, sales variance replaces demand variance in each expression and in addition sales variance will not be independent of price.

$$\text{(18)} \quad \underset{p_1,\ldots,p_n,Q}{\text{Maximize}}\ E[\pi] = \sum_{i=1}^{n} \int_{-\bar{D}_i(p_i)}^{Q-\bar{D}_i(p_i)} g_i(u_i)[\bar{D}_i(p_i) + u_i]du_i + \sum_{i=1}^{n} (p_i - b)Q\{1 - G_i[Q - \bar{D}_i(p_i)]\} - \beta Q$$

subject to

$$\text{(19)} \qquad Var\ [\pi] \leq \bar{\sigma}^2$$

$$\text{(20)} \quad p_i \geq 0, \quad i = 1, \ldots, n\ Q \geq 0$$

The necessary condition for optimal prices from (18)–(20) consists of equation (6) plus a second term

$$\text{(21)} \qquad -\lambda \frac{\partial\, Var\ [\pi]}{\partial p_i} \qquad i = 1, \ldots, n$$

where $\lambda \geq 0$ represents the marginal profit for a small change in total variance at the constraint boundary. The term in (21) measures the economic cost of the marginal risk contributed by the ith group of customers. An optimal structure of prices under risk will obviously contain differences which reflect differences in the contributions each class of customers makes to total risk. Those with stable demands may contribute little or nothing to total risk and thus their prices will in general be lower, while those with unstable demand will be required to bear a risk premium which is essentially the insurance premium they must be willing to pay for exercising a demand option on capacity. The foregoing provides a partial explanation for the practice of offering special terms to "steady" customers, i.e., ones with low demand variance.

The presence of a risk-efficient aspect to pricing means one must proceed with special caution in attempting to infer whether price discrimination is present by a comparison of prices and the concept of marginal cost reckoned in traditional riskless setting.

V. Conclusion

This paper has presented a set of results for a monopolist's optimal choice of pricing and investment policy under both riskless and stochastic conditions. The riskless results were, of course, only included for completeness. Although the stochastic results are based on the additive error model, the results obtained include the special case suggested by Brown and Johnson, but do not require specific assumptions about either the demand function or about the probability distribution for the random component.

An optimal pricing/capacity choice policy will generally imply capacity is *not* fully utilized all of the time. This is not inefficiency, but rather is the optimal response on the part of a firm which seeks to limit risks. While the argument suggested by Averch and Johnson implies excess investment may be sought by a regulated firm for less pristine motives, *apparent* overengineering or *apparent* excessive investment may be simply an optimal economic decision when the firm must meet high reliability standards for the service it provides. Such a decision is not necessarily inefficient once one realizes that two efficiency conditions must be satisfied—one stemming from the traditional riskless economic setting and a risk-efficiency condition which emerges only under uncertainty.

The concept of risk efficiency in pricing and output choice is more general than the demand uncertainty case analyzed here. When applied to demand uncertainty where only total revenue (or net operating revenue) is the random component of profits, several interesting interpretations emerge for the role of prices in altering risk. An economic interpretation of two earlier results comparing riskless and risky pricing rules are seen to follow naturally from introducing the notion of risk-efficient pricing. In addition, one finds that the stochastic nature of demand may provide a rationale for price discrimination which is quite different from the traditional riskless one, since altering the "mix" of prices over classes of customers can change profit and/or excess demand risk.

REFERENCES

K. J. Arrow, "Optimal Capital Policy, the Cost of Capital, and Myopic Decision Rules," *Institute of Statistical Mathematics Annals*, 1964, *16*, 21–30.

H. Averch and L. L. Johnson, "Behavior of the Firm Under Regulatory Constraint," *Amer. Econ. Rev.*, Dec. 1962, *52*, 1052–69.

E. E. Bailey, "Peak-Load Pricing Under Regulatory Constraint," *J. Polit. Econ.*, July/Aug. 1972, *80*, 662–79.

M. Boiteux, "Peak-Load Pricing," in J. R. Nelson, ed., *Marginal Cost Pricing in Practice*, Englewood Cliffs 1964.

G. Brown and M. B. Johnson, "Public Utility Pricing and Output Under Risk," *Amer. Econ. Rev.*, Mar. 1969, *59*, 119–28.

——— and ———, "Public Utility Pricing and Output Under Risk: Reply," *Amer. Econ. Rev.*, June 1970, *60*, 489–90.

P. Dhrymes, "On the Theory of the Monopolistic Multiproduct Firm Under Uncertainty," *Int. Econ. Rev.*, Sept. 1964, *5*, 239–57.

D. W. Jorgenson, "Capital Theory and Investment Behavior," *Amer. Econ. Rev. Proc.*, May 1963, *53*, 247–59.

A. E. Kahn, *The Economics of Regulation*, Vol. I, New York 1970.

S. Karlin and C. Carr, "Prices and Optimal Inventory Policy," in K. J. Arrow, S. Karlin, and H. Scarf, eds., *Studies in Applied Probability and Management Science*, Stanford 1962.

H. E. Leland, "Theory of the Firm Facing Uncertain Demand," *Amer. Econ. Rev.*, June 1972, *62*, 278–91.

E. Mills, "Uncertainty and Price Theory," *Quart. J. Econ.*, Feb. 1959, *73*, 116–30.

H. Mohring, "The Peak-Load Pricing Problem with Increasing Returns and Pricing Constraints," *Amer. Econ. Rev.*, Sept. 1970, *60*, 693–705.

J. N. Rosse, "Product Quality and Regulatory Constraint," Studies in Industry Economics, no. 22, Stanford Univ., Nov. 1972.

P. O. Steiner, "Peak-Loads and Efficient Pricing," *Quart. J. Econ.*, Nov. 1957, *71*, 585–610.

R. Turvey, "Public Utility Pricing and Output Under Risk: Comment," *Amer. Econ. Rev.*, June 1970, *60*, 485–86.

———, "Peak-Loading Pricing," *J. Polit. Econ.*, Jan./Feb. 1968, *76*, 101–13.

M. L. Visscher, "Welfare-Maximizing Price and Output with Stochastic Demand: Comment," *Amer. Econ. Rev.*, Mar. 1973, *63*, 224–29.

O. E. Williamson, "Peak-Load Pricing and Optimal Capacity Under Indivisibility Constraints," *Amer. Econ. Rev.*, Sept. 1966, *56*, 810–27.

Dynamic Monopsony and Structural Change

By MARTIN NEIL BAILY*

This paper examines at the microeconomic level the assumption that the rate of change of the labor supply to a firm depends on the wage paid by the firm (relative to the rest of the economy). The firm is, therefore, a dynamic monopsonist. First, the behavior of a single firm is considered and its response in terms of the optimal wage path to different product prices. Second, an industry of N such firms is analyzed with a static product demand, and the convergence of such an industry to stationary equilibrium is shown. Third, the assumptions are changed so that instead of operating in a static environment, the industry experiences changing demand, changing labor productivity, and a rising wage in the rest of the economy. In the growing environment the model predicts a pattern of wage differentials in response to steady-state trends. An additional feature of the model is that it is assumed that there is some minimum wage such that if the firms offer less than this all its workers quit. This restriction introduces the possibility of layoffs as an optimal policy by the firm. This is of interest since the monopsony assumption has been criticized (for example by Herschel Grossman) on the grounds that layoffs are excluded.

The idea of dynamic monopsony is not new. Earlier versions of this model were developed some time ago based in part on verbal discussions of the idea from a number of sources (for example, Paul Samuelson's text, p. 554). Edmund Phelps and Sidney Winter, Dale Mortensen, and other search theorists have used dynamic monopoly and monopsony assumptions to examine questions of inflation and unemployment. It should be emphasized that the model presented here takes macro variables such as the overall unemployment rate and the price level as fixed.

I. The Monopsonistic Firm

Consider a single firm which has an initial allocation of labor $L^s(0)>0$. The labor supply conditions it faces are given by:

$$\frac{DL^s}{L^s} = g(w) \tag{1}$$

where $g' > 0,\ g'' < 0,$

$$g(w) \rightarrow -\infty \text{ as } w \rightarrow w_{min} > 0$$

w is the money wage offered by the firm, but since we are taking overall conditions in the economy as constant, we can take w as the wage offer by this firm relative to the rest of the economy.[1] The specification has a certain amount of inherent plausibility, given the existence of mobility costs in the labor market. If the wage is low not all workers quit right away, but over time quits exceed hirings and the available labor supply declines. If the wage is high, quits are low and new workers are attracted to the firm. Recent literature has emphasized the informational and search costs involved in labor mobility.[2] As well as this, one would expect that even if the pattern of wages were known, the costs and possible disutility of relocation plus the cost of learning the new job would mean that the response to wage differentials would occur over time.

---

* Assistant professor of economics, Yale University. I would like to thank F. M. Fisher, R. M. Solow, P. A. Diamond, M. A. Baily, W. C. Brainard, the members of the Harvard-MIT mathematical economics seminar, and the managing editor and referee for many helpful comments. I retain responsibility for error. This is a considerably revised version of part of my 1972 MIT doctoral dissertation. Financial support from the Canada Council is acknowledged.

[1] Later we allow the general level of wages in the rest of the economy to be rising. This makes g shift over time.

It is specified that the *rate* of growth or decline of the labor supply depends upon the wage. This means that for a given w, a large firm will lose or gain a proportionally large number of workers. It seems plausible that when w is low then a proportion of the work force will decide to quit during any given period. It is also likely that the larger the firm the greater the number of workers who will hear about jobs from workers already employed or who will just turn up to see what opportunities are available. It would be wrong to disguise, however, that there is mathematical expediency involved also. The analysis could be made with a more general form, but it would make things much more awkward.

A more important limitation of the form used is that it embodies static expectations. Workers respond to the current wage. Parts of this paper will be concerned with cases where the static expectations are fulfilled. Whenever we consider a path of w which varies over time then the assumption becomes stronger. One can take the assumption of concavity of $g(w)$ as a proxy for a certain type of worker expectation. If the wage is increased from a low starting point, workers think the increase is more likely to persist than if it is increased from a high starting point; hence decreasing returns to successive wage increases. The concavity assumption is important for sufficiency of an optimum and for stability.

The production conditions for the firm are described by a constant returns production function.

$$X = F(K, L) = Lf(k), \qquad k = K/L, \qquad \frac{\partial F}{\partial L} > 0, \qquad \frac{\partial F}{\partial K} > 0, \qquad F \text{ concave}, \qquad F(K, 0) = 0 \tag{2}$$

where X is the output of the firm, K is the amount of capital rented, and L is the number of man-hours of labor services used.[3] In order to focus attention on the implications of the dynamic monopsony, some simple assumptions are made about the product and capital markets. The firm is allowed to rent as much capital as it wishes at a constant rental rate ρ. Therefore K is freely adjustable. On the product side it is assumed that the firm can be assumed, at least as an approximation, to be a price taker.[4] While alternative assumptions could well be made, the above has the advantage that we can examine in a straightforward way how a firm adjusts its wage in response to different values of the price it faces.

The firm, therefore, chooses values of its control variables capital, labor, and the wage to maximize the present value of its profits:

$$\max_{K,L,w} \int_0^\infty e^{-rt}[p(t)F(K, L) - wL - \rho K]dt \tag{3}$$

$$\text{subject to} \quad L \leq L^s, \quad DL^s = g(w)L^s, \quad L^s(0) \text{ given}, \quad L \geq 0, \quad K \geq 0$$

---

[2] See, for example, George Stigler, John McCall, and in particular Mortensen who gives a detailed search model that leads to a labor supply like (1). Other types of mobility cost and the effect of workers forming expectations of future wages are discussed by the author in an earlier paper.

[3] There is a complication that has been ignored. The firm can raise the number of hours worked per employee by paying overtime wage rates. A standard work week per employee is assumed here.

[4] The equilibrium of the industry including the possible entry and exit of firms is discussed in Section III.

The only state variable is L^s, the supply of labor to the firm. If on the optimal path we have $L<L^s$ then the firm is not using all the labor which offers itself. This corresponds to a layoff or a complete shutdown. This possibility is dealt with below. Consider first, therefore, the case $L=L^s$, $K>0$, and $L>0$. This simplifies the problem. The firm sets the wage and the amount of capital it wants and accepts the available labor supply. The necessary maximizing conditions are:[5]

$$(4)\qquad p(t)\frac{\partial F}{\partial K}=\rho$$

$$(5)\qquad -L+\lambda g'(w)L=0 \quad \text{or} \quad \lambda=1/g'(w)$$

$$(6)\qquad D\lambda=r\lambda-p(t)\frac{\partial F}{\partial L}+w-\lambda g(w)$$

plus the transversality conditions:

$$(7)\qquad \lim_{t\to\infty} e^{-rt}\lambda(t)\geq 0 \quad \lim_{t\to\infty} e^{-rt}\lambda(t)L(t)=0$$

and $L(t)$ must also satisfy the labor supply condition (1). Equation (4) says that the value of the marginal product of capital is always equated to the rental rate on capital. This is not surprising given the assumption of costless capital adjustment. The conditions determining the wage path are necessarily somewhat more complicated. The term $g'(w)$ is the marginal effect of a wage change in inducing more rapid growth of the available labor supply. The larger is $g'(w)$, the more responsive is the labor supply. The inverse of $g'(w)$ is equal to λ (equation (5)) which is the shadow price of the constraint, i.e., the cost to the firm of adding one more worker. Equation (6) then says that the cost λ of a marginal worker is just equal to the benefit to the firm of the marginal worker. To see this, multiply (6) through by the factor e^{-rt} and then integrate to obtain:

$$(8)\qquad \int_0^\infty e^{-rt}[r\lambda-D\lambda]dt=\lambda(0)=\int_0^\infty e^{-rt}\left[p\frac{\partial F}{\partial L}-w+g(w)\lambda(t)\right]dt$$

Hence the wage is set so that the shadow price λ of an extra worker now is just equal to the present value of the benefit—made up of two parts. First, there is the excess of the marginal product over the wage. Second, there is a term that is zero whenever the wage is set to maintain a constant labor supply $[DL^s/L^s=g(w)]$, is positive whenever the labor force is expanded, and negative whenever there is a contraction. If it is planned to expand the labor force in the future, then an extra worker hired now is valuable because he makes it easier to hire future workers (for example, by telling potential workers about the high wage available). If it is planned to contract the labor force in the future, then we must make a subtraction from the present value of the excess of product over wage because the worker hired now may be among those who quit later, so his product does not accrue for the full time.[6] Since by assumption, $g(w)$ is concave and $F(K, L)$ is concave, the necessary conditions are also sufficient for a maximum, and a path satisfying these conditions will be optimal for the problem as originally stated in (3) unless a shutdown or layoff is better.

II. Properties of the Optimal Path

Since the production function is constant returns and well-behaved, the marginal products depend only on the capital-labor ratio k and are monotonic functions of this argument. Equation (4) therefore determines $k(t)$ as a function of $p(t)$ for a given rental rate ρ. If we then substitute (4) and (5) into (6), we obtain:[7]

---

[5] The standard reference is L. S. Pontryagin et al. The exposition in Kenneth Arrow and Mordecai Kurz is also helpful. Some interpretation of the conditions is given here for the sake of clarity, but it is kept fairly brief because at this point much of the ground is familiar from Mortensen, pp. 184–86.

[6] Strictly speaking, since the variables are all continuous and nonstochastic, the mathematics say that some fraction of any workers hired now will quit later.

$$(9)\quad p(t)\frac{\partial F}{\partial L} = h(p(t)) = w + \frac{r}{g'(w)} + \frac{g''(w)Dw}{[g'(w)]^2} - \frac{g(w)}{g'(w)}$$

The function h is the value of the marginal product of labor evaluated at the optimal capital-labor ratio. It depends only on current price (since k is freely adjustable) and is a monotonic increasing function of this argument. This is easy to see since when price is high, (4) tells us that k is also high and so the marginal product of labor is high. It is worth keeping the definition of $h(p(t))$ in mind for the subsequent discussion.

Some properties of the optimal wage path determined by (9) will now be derived. Proofs of the results are in the Appendix. In general, the wage set at any instant depends upon the future path of the product price. If the firm has perfect foresight there is no problem, but more realistically we think of $p(t)$ as the expected path of the price. This, of course, still abstracts from problems of attitude towards risk. The point at which the function $g(w)$ crosses the axis is an important reference point for the path. Define $\bar{w}$ by $g(\bar{w})=0$, *so* $\bar{w}$ *is the wage the firm must pay in order to just retain the same size work force.* It is tempting to assume immediately that $\bar{w}$ must be the wage prevailing in the rest of the economy (if the total labor force is constant) for then if all firms pay the same wage none of them gains or loses workers. The general equilibrium properties of models with search or labor mobility are quite tricky, however, and so, depending upon the nature of the mobility process and on how tight or loose the overall labor market conditions are, $\bar{w}$ may in general be above or below the economy average wage.[8] This model is of partial equilibrium, however, and presumably *there is some overall condition of the labor market such that* $\bar{w}$ *is the economy average wage.* Although it is not essential to assume this neutral case, it is convenient to do so when interpreting the results.

The first feature of the firm's decision we consider is under what circumstances will the firm be content to keep its labor force constant.

PROPERTY 1. *There is a unique price* $\bar{p}$ *that is consistent with stationary* w, L, *and* K *on the optimal path.* This property follows from (9) and the fact that $h(p)$ and $g(w)$ are monotonic. Clearly $\bar{p}$ the price which causes the firm to keep its labor force constant is going to be determined by the size of $\bar{w}$ the necessary wage to achieve this constancy. In fact $\bar{p}$ and $\bar{w}$ are related by:

$$(10)\qquad h(\bar{p}) = \bar{w} + r/g'(\bar{w})$$

The stationary solution is similar to the result one would obtain from a static monopsony assumption. If there were a static rising supply curve of labor then we would have the analogous equation:

$$(11)\qquad h(p) = w + w/E$$

where E is the elasticity of labor supply. The implicit elasticity of labor supply in (10) is $\bar{w}g'(\bar{w})/r$.

If the price is constant over time but not equal to $\bar{p}$ then the optimal path will be to pay a constant wage provided the price lies within a certain interval. The firm cannot pay less than w_{min} without losing all its workers. There is a price, p_{min}, such that any constant price below this will cause the firm to shut down rather than incur the loss that would follow from paying any wage above w_{min}. If

---

[7] Since the production function is assumed to have constant returns and the labor supply condition determines the rate of change of L, equation (9) is independent of $L(0)$. This means the optimal wage path is independent of the initial condition.

[8] This point is made with respect to Phelps and Winter by William Nordhaus. In a general equilibrium context it is presumably not even valid to consider just the first moment of the distribution of wage offers by other firms, but here this distribution is taken as given.

$$h(p_{min}) = w_{min} \tag{12}$$

where w_{min} is given in (1), then any constant price below p_{min} must lead to loss or to shutdown. There is a maximum price set by the necessary conditions. If the price is too high, the integral (8) does not converge.[9] We require $r > g$, so that p_{max} is given by:

$$h(p_{max}) = w_M \tag{13}$$

where $w_M = g^{-1}(r)$

PROPERTY 2. *For any constant price* $\tilde{p}$ *such that* $p_{min} < \tilde{p} < p_{max}$ *there is a unique optimal wage path. This path is to set a constant wage* $\tilde{w}(\tilde{p})$ *where* $\tilde{w}$ *and* $\tilde{p}$ *are related by:*

$$h(\tilde{p}) = \tilde{w} + [r - g(\tilde{w})]/g'(\tilde{w}) \tag{14}$$

Properties 1 and 2 show that the firm responds in a natural way to the product price. Equation (14) shows a high product price ($\tilde{p} > \bar{p}$) will cause the firm to expand and the opposite for a low price. The difference between $\bar{p}$ the stationary equilibrium price and $\tilde{p}$ is shown to be directly related to the speed of adjustment of the firm, a property of adjustment paths that is often simply assumed. A property of more general price paths that will be useful for showing industry equilibrium is:

PROPERTY 3. (a) *If the price* $p(t)$ *is always above (below)* $\tilde{p}$ *then the optimal wage path* $w^*(t)$ *is always above (below)* $\tilde{w}$ *from (14).* (b) *If* $p(t)$ *converges from above (below) to* $\tilde{p}$ *where* $p_{min} < \tilde{p} < p_{max}$ *then the optimal path converges from above (below) to* $\tilde{w}$.

We have excluded the possibility of shutdown, but stated under what conditions it will occur, i.e., when the firm cannot avoid a loss. For a general price path (not necessarily constant), the shutdown could occur initially or, if price is decreasing over time, the firm might operate for a while and then shut down later. The possibility of layoffs, where some but not all workers are released, is more subtle. The only way this could occur would be if the firm were just indifferent between shutting down and continuing to operate, i.e., if the present value of the firm were zero when the optimal wage path was followed. Looking only at a single firm, it would seem a very remote possibility that price would just follow such a path. This is not so, as we shall see in the next section.

III. Industry Equilibrium

We now consider the industry which is made up of N firms like the one we have been discussing. The production function is assumed the same for all firms, and they all face the same product price and the same form of equations (4) and (9). This means that they will all set the same wage path and the same capital-labor ratio.[10] It is, of course, a simplifying device to consider a group of essentially identical firms, but it is a useful one. We can easily determine total industry supply and show that for any given demand function, the industry will converge to a stationary equilibrium. Total supply is given by:

---

[9] If the price remains above p_{max} for all time then the present value of the firm goes to infinity. See also the discussion of entry.

[10] One could allow for differences in efficiency or capital vintage across firms. This would complicate things considerably, however, because there would be a different wage for each firm. The industry output and price would depend upon the initial allocation of labor between firms.

$$
\begin{aligned}
(15)\quad X^s &= \sum_{1}^{N} F(K_i, L_i) \\
&= \sum L_i(\mathrm{t})f(k_i(\mathrm{t})) = \sum L_i(\mathrm{t})f(k(\mathrm{t})) \\
&= L(\mathrm{t})f(k(\mathrm{t})) \quad \text{where } L(\mathrm{t}) = \sum L_i(\mathrm{t})
\end{aligned}
$$

where L_i and K_i are now the employment and capital of the ith firm, L is the total industry employment, and $k(t)$ is the capital-labor ratio, common to all firms.

Industry demand is given by:

$$(16)\qquad X^d = Q(p(\mathrm{t})), \qquad Q' < 0$$

and it is assumed that there is always equilibrium in the product market so that:

$$(17)\qquad Q(p(\mathrm{t})) = L(\mathrm{t})f(k(\mathrm{t}))$$

Equations (1), (4), (9), and (17) will determine the path of price, wage, employment, and capital in the industry from any initial starting point $L(0)$. First, is there a stationary point for the industry? Second, will the industry converge to this point? If we constrain the demand curve by the condition

$$(18)\qquad Q(\bar{p}) > 0$$

where $\bar{p}$ is defined in Property 1, then continued production is profitable. There will then be a stationary solution for which $w = \bar{w}$ and $p = \bar{p}$, and in addition:

$$
\begin{aligned}
(19)\quad & k = \bar{k} \text{ from } \bar{p}f'(\bar{k}) = \rho \\
\text{and}\quad & L = \bar{L} \text{ from } Q(\bar{p}) = \bar{L}f(\bar{k})
\end{aligned}
$$

In general we would not expect the initial industry labor force to be just equal to the value $\bar{L}$ in (19). If $L(0) \neq \bar{L}$ we have the following:

THEOREM 1: *If all firms set the optimal wage path, then starting from any initial condition* $L(0) > 0$, *the paths of* $L(t)$, $p(t)$, $k(t)$, *and* $w(t)$ *will all converge to* $\bar{L}$, $\bar{p}$, $\bar{k}$, *and* $\bar{w}$.

PROOF:

Equations (4) and (17) can be solved for $p(\mathrm{t})$ and $k(\mathrm{t})$ in terms of $L(\mathrm{t})$: $L(\mathrm{t}) \lessgtr \bar{L} \Rightarrow p(\mathrm{t}) \gtrless \bar{p}$ and $k(\mathrm{t}) \gtrless \bar{k}$. If at any time τ, $L(\tau) = \bar{L}$, then $p(\tau) = \bar{p}$ and all firms would be following an optimal policy if they were to set $w(\mathrm{t}) = \bar{w}$ for all $\mathrm{t} \geq \tau$. This means that if the initial industry labor force is less than (greater than) the equilibrium value then it will never cross and become greater than (less than) the equilibrium level. Hence $L(0) \lessgtr \bar{L} \Rightarrow p(\mathrm{t}) \gtreqless \bar{p}$ for all $\mathrm{t} \geq 0$. We can therefore apply Property 3 to give:

$$(20)\qquad L(\mathrm{t}) \gtreqless \bar{L} \Rightarrow w(\mathrm{t}) \lesseqgtr \bar{w}$$

A simple Lyapunov function $V = 1/2(L(\mathrm{t}) - \bar{L})^2$ gives $DV = (L(\mathrm{t}) - \bar{L})DL$ and this ensures $L(\mathrm{t})$ converges to $\bar{L}$. Equations (4) and (17) then imply $p(\mathrm{t})$ and $k(\mathrm{t})$ converge to $\bar{p}$ and $\bar{k}$, and Property 3 implies $w(\mathrm{t})$ converges to $\bar{w}$.

A similarity was shown earlier between static monopsony and the stationary equilibrium in this model in that the wage is below the value of the marginal product. We now consider whether or not this means there are pure profits that would attract new entrants to the industry. This model is not the same as the case of a single monopsonist, however. The ability to extract a quasi rent from each worker derives from the monopsony power held over existing workers. A new entrant does not have any workers, but would have to acquire them by paying a wage premium. In fact, the quasi rent is not a pure profit. It is a return to the investment made during the acquisition of a labor force.

We do require a more complete specification of the labor supply condition which, as it stands, is not applicable to an entering firm. While space does not permit a full formal analysis it is not hard to see what is necessary. We now assume that *a new firm must incur a start-up cost, which* *includes the hiring of an initial minimum labor force,* that is at least as great as the cost to an existing firm of an equal expansion in its labor force. This seems reasonable enough since the kind of informational lags and reluctance to change jobs that created the labor supply conditions to existing firms will apply at least as much to potential entrants. Once established a new firm can expand according to equation (1).

Consider, therefore, a firm in equilibrium. Take equation (10) and express it more instructively as $h(\bar{p})=\bar{p}\partial F/\partial L=\bar{w}+r\lambda(\bar{w})$ using equation (5). We have seen that λ is the shadow price of a worker, the cost to the firm of acquiring an extra worker and his value to the firm. The worker is acquired by a kind of investment, by paying a wage premium over time. Hence the above condition states that the marginal product is equal to the wage cost of the worker plus the interest cost $r\lambda$ of an asset of value λ. Therefore the firm is not earning a pure profit, but only a normal rate of return r on the value of its past investment in obtaining a labor force. Since the marginal product of capital is always set equal to the rental rate on capital, the two conditions imply that the firms are operating where price just covers average total cost. Total cost includes the interest cost of the investment in workers. The potential entrant, by assumption, must pay at least λ for each worker so it will see only a normal or less than normal rate of return and will not be encouraged to enter.

The above argument will hold, of course, even more if price is below $\bar{p}$. Such a fall in demand would impose a capital loss on existing firms as the value of their acquired stock of workers falls. That is why they start to disinvest or reduce their labor force. In fact there is some L_{max} such that if the initial labor force were greater than this, price would be so low that firms would have a negative present value for all wage paths above w_{min}. The value of a worker has dropped to zero and firms will start to lay off workers. The price will then rise until the point where losses can be avoided and firms will then stop laying off.[11] Beyond this point the shadow price (and hence value to the firm) of each worker is again positive.

If on the other hand there is an increase in demand, it is natural to think that at some point when price is high enough, entry would occur. The only consequence for the model of such entry is that prices above this point would not be observed since they would generate entry and lower the price again. If the start-up costs were very large we could observe prices up to the p_{max} from equation (13). If the costs are smaller then there will be an effective maximum p_{max} between $\bar{p}$ and the solution of (13). If we redefine p_{max} in this way, it gives an economic condition more symmetric with p_{min} than the arbitrary mathematical condition (13).[12] Within the range p_{min} to the new p_{max} we have the interior solution given above. Outside this range we have shutdown or entry.

IV. Product Demand and the Wage

The framework in which the industry has been assumed to operate is purely static. We now assume that the overall economy is growing. There is labor-augmenting technical progress at a rate δ and population growth at a rate n. Consider the following form for the labor supply function:

$$\text{(21)} \quad \frac{DL^s}{L^s} = g(w,\, t) = g(we^{-\delta t})$$

and $\quad g(\bar{w}) = n$

---

[11] The firms stop laying off workers. They are still reducing the size of their work forces.

[12] Depending upon a taste for detail one could distinguish a p_{max} that would stimulate entry if it were sustained (as in the next section), and a price that was higher, but expected to fall.

The point of reference of the labor supply function has been shifted from zero to n to take account of the effect of the overall growth of the labor force. In addition the function shifts over time because technical progress is raising wages throughout the economy. The firms experience the benefits of the technical progress in terms of increased productivity as well as the cost of rising wages elsewhere. The maximization problem *for each firm* (excluding layoffs) is therefore given by:

$$(22)\qquad \max_{w_i, K_i} \int_0^\infty e^{-rt}[pF(K_i, L_i e^{\delta t}) - w_i L_i - \rho K_i]dt$$

$$\text{subject to}\quad \frac{DL_i}{L_i} = g(w_i e^{-\delta t})$$

If we define $y_i(t) = w_i e^{-\delta t}$ and define $k_i = K_i / e^{\delta t} L_i$ then (22) becomes:

$$(23)\qquad \max_{y_i, k_i} \int_0^\infty e^{-rt} e^{\delta t}[pL_i f(k_i) - y_i L_i - \rho k_i L_i]dt$$

$$\text{subject to}\quad \frac{DL_i}{L_i} = g(y_i)$$

This problem is exactly the same as the original one except $r-\delta$ replaces r. The same stability and convergence properties will therefore apply, and all firms will set a common wage and capital-labor ratio.

Population growth and technical progress mean that the overall economy is growing. We will assume that the rate of growth in the economy is at a constant rate $\alpha = \delta + n$. The rising income level will affect the demand for the industry's product. If we let the income elasticity of demand for the product be a constant m, then we have:

$$(24)\qquad X^d = Q(p)Y^m = Q(p)[Y_0 e^{\alpha t}]^m = Q(p)e^{m\alpha t}$$

where Y_0 is absorbed into the function Q. Total industry supply is given by

$$(25)\qquad X^s = \sum_1^N L_i e^{\delta t} f(k) = Le^{\delta t} f(k)$$

So that putting these together gives:

$$(26)\qquad Q(p)e^{m\alpha t} = Le^{\delta t} f(k)$$

The stationary solution of the previous section is now transformed into a steady-state solution with the properties:

$$(27)\qquad \begin{aligned} (DL/L)^* &= m\alpha - \delta \\ &= m(\delta + n) - \delta \\ \overset{*}{w} &= \overset{*}{w}_0 e^{\delta t} \\ g(\overset{*}{w}_0) &= m(\delta + n) - \delta \\ \overset{*}{p}[f(\overset{*}{k}) - \overset{*}{k} f'(\overset{*}{k})] &= \overset{*}{w}_0 + \frac{r - m(\delta + n)}{g'(\overset{*}{w}_0)} \\ \overset{*}{p} f'(\overset{*}{k}) &= \rho \\ Q(\overset{*}{p}) &= \overset{*}{L}_0 f(\overset{*}{k}) \end{aligned}$$

Such a solution will exist and the firms will all be following their optimal policies provided: (i) $L_0 = L^*$, (ii) initial demand for the product is large enough that $Q(p^*) > 0$, and (iii) that $r > m(\delta + n)$ so that the transversality conditions are satisfied. If $L_0 \neq L_0^*$ then Property 3 ensures that the industry will converge to the steady-state path.

THEOREM 2: *Starting from any initial industry labor force* (L_0), *then if all firms set the optimal wage and output path, the product price* $p(t)$, *the industry labor force* $L(t)$, *the capital-effective labor ratio* $k(t)$, *and the wage* $w(t)$ *will all converge to their equilibrium values* p^*, $L_0^* e^{[m(\delta+n)-\delta]t}$, k^*, *and* w^* (*from* (27)).

PROOF.

The theorem can be shown by the same method as that of Theorem 1.

On the steady-state path there is a simple relation between the product demand and whether or not the industry is expanding, high wage or contracting, low wage. Specifically:

$$(28) \qquad w^* \gtreqless \bar{w}$$

and hence $(DL/L)^* \gtreqless n$ as $m \gtreqless 1$

This result can easily be seen from (27) where m is the income elasticity of demand for the product. In an expanding economy, therefore, an industry that is experiencing relatively increasing demand for its product as income rises ($m>1$) will be a high wage industry. The interaction between product demand and the wage is hence of quite a different kind in the context of a growing economy. Discussions of the structure of wages usually see high or low wages as a disequilibrium response to unanticipated shifts in demand. Indeed the earlier results of this model were quite consistent with this view. In the growing economy, however, *wage differentials will occur as a natural response to the changing structure of demand.* Of course if the initial labor force is not L_0^*, then there will be an adjustment with the wage above or below its steady-state value.

The changing structure of demand is only a part of the story. So far, technical progress played only a very minor role in the results because the increasing productivity of labor was balanced by the rising level of wages throughout the economy.

V. Technological Change, Demand, and the Wage

We keep most of the framework of the previous section with the following modifications. The rate of technical change experienced by firms in this industry is γ which may be different from the rate for the economy as a whole δ.[13] It is a sufficient condition for a steady state to exist if we assume a Cobb-Douglas production function and a constant price elasticity of demand. Industry demand is then:

$$(29) \qquad X^d = ap^{-b}Y^m = ap^{-b}e^{mat}$$

where a and b are positive constants and Y_0^m is absorbed into a. The industry supply is given by:

$$(30) \qquad X^s = [Le^{\gamma t}]k^{\beta}, \qquad k = K/Le^{\gamma t}, \qquad 0<\beta<1$$

The steady-state solution is characterized by:

$$\begin{aligned}(31) \quad Dp/p &= (1-\beta)(\delta-\gamma)\\ Dk/k &= (\delta-\gamma)\\ DL/L = \theta &= (\delta-\gamma)(1-b)(1-\beta)\\ &\quad + mn + (m-1)\delta\\ &= g(\hat{w}_0) \quad \text{and} \quad \hat{w} = \hat{w}_0 e^{\delta t}\end{aligned}$$

There is also a set of initial conditions and we must have $r>\theta+\delta$ to satisfy the transversality conditions. The rate of technical change relative to the rest of the economy $(\delta-\gamma)$ determines the direction of changes in the steady-state price and capital-labor ratio. A rapid rate of technical change $(\gamma>\delta)$ means that labor costs are falling and this would mean a falling price of this industry's product. Looking at the capital-labor ratio, one can recall that the value of the marginal product of capital is always equal to the rental rate—assumed constant. A falling price lowers the value of the marginal product which must be offset by adjusting the capital-labor ratio. The above arguments hold in reverse for slow technical change.

The conditions that lead to a high or low wage are obtained from the third part of (31):

$$(32) \quad \hat{w} \gtreqless \bar{w} \quad \text{as} \quad \alpha(m-1) + (\delta-\gamma)(1-b)(1-\beta) \gtreqless 0$$

The first term of this expression is the same as the previous section; $\alpha=n+\delta$ is the rate of growth of income in the economy. This enters because for any income elasticity of demand, a different rate of growth leads to more or less rapid changes in the structure of demand. The term is positive or negative as m is greater or less than unity just as before. The effect of technical change is interesting in part because it is ambiguous. *Whether or not industry-specific technical change helps or hurts workers, i.e., leads to a high or low wage, depends upon the price elasticity of demand for the product.*

---

[13] This industry is small relative to the economy so that its rate of technical progress has a very small effect on the rate for the economy.

Rapid technical progress means that any given output can be produced with less labor.[14] The firms can economize on labor and keep the wage down. On the other hand, rapid technical progress allows the price to fall and this increases the quantity demanded. This effect tends to make the industry expansionary and high wage. Which of these two forces predominates depends on how elastic demand is. If the fall in price is a strong stimulus to demand ($b>1$), then rapid technical change will imply a higher wage than would be true for $\gamma=\delta$. If demand is inelastic ($b<1$), then rapid technical change implies economizing on labor and a lower wage.

VI. The Structure of Wages

Bearing in mind that caution needs to be exercised in applying a simplified model to empirical relations, we can nevertheless use the results when studying the pattern of relative wages. The dynamic monopsony specification *provides a framework to help understand that part of the relative wage structure associated with different growth rates, demand conditions, and rates of technological change.* Static supply and demand models of the labor market have to rely on *ad hoc* specification in order to explain why fast growing firms or industries tend to be high wage payers. Such models also give little information on the nature of the time path of the wage.

There is a very large literature on wage differentials, much of which focuses on skill and education differences. There is also a considerable literature on wages and mobility,[15] which gives some support to the underlying assumption of this model. Various incentive effects including a high relative wage reduce the quit rate and encourage hirings. It is common to include a measure of profitability to explain wage differentials.[16] One rationale for this is that workers ask for higher wages when profits are high. While this may be true, there is also a clear prediction from this model that a high product price that will imply high profits, will also imply a high wage set by the firm as a profit-maximizing path. The differences between neoclassical and institutional approaches to wage setting can hopefully be more easily clarified and resolved by an explicit dynamic profit-maximizing analysis.[17]

While many of the results of empirical studies are consistent with the model of this paper, there is not, to my knowledge, any paper that tests directly the predicted relations between wage and demand. One possible example that seems to fit the predicted result is agriculture, at least for an extended period up to 1972. The income elasticity of demand is low; the rate of technical progress has been very high; the price elasticity is low. Thus there is an unambiguous result from (32) implying a low wage declining industry. The limitations of steady-state results are indicated by the change in conditions experienced by agriculture in 1973–74 with the sudden spurt in foreign demand. Nevertheless, there were clear trends over a number of years which apparently fit this model.

---

[14] Part 2 of equation (31) shows that the capital-labor ratio is falling when $\gamma>\delta$ so that the marginal product of labor is falling. Also price is falling when $\gamma>\delta$ so that the value of the marginal product is certainly falling. Hence for constant output, the firm would economize on labor.

[15] See, for example, OECD, Lloyd Ulman, John Burton and John Parker, and references given by these authors.

[16] See, for example, Stuart Schweitzer.

[17] Ulman gives a discussion of alternative approaches and references.

APPENDIX

In this Appendix proofs are given of Properties 1–3 of the optimal path stated in Section II.

PROPERTY 1. *There is a unique price $\bar{p}$ that is consistent with stationary w, L, and K on the optimal path.*

PROOF:

If L is stationary then $DL=0$ and $w=\bar{w}$ so that equation (9) becomes:

$$h[p(\mathrm{t})] = \bar{w} + r/g'(\bar{w}) \tag{A1}$$

Given that, as we saw earlier, h is a monotonic function of p, (A1) defines a unique solution $p(\mathrm{t})=\bar{p}$. If $p=\bar{p}$, $w=\bar{w}$, and k satisfies equation (4) then all the necessary conditions of the problem are satisfied. In addition, F and g are concave and the transversality conditions are satisfied. Hence if $p(\mathrm{t})=\bar{p}$ for all $\mathrm{t}\geq 0$, we have found an optimal solution. The proof of Property 2 will show it is *the* optimal solution.

PROPERTY 2. *For any constant price $\tilde{p}$ such that $p_{min}<\tilde{p}<p_{max}$ there is a unique optimal wage path. This path is to set a constant wage $\bar{w}(\tilde{p})$ where $\tilde{p}$ and $\bar{w}$ are related by:*

$$h(\tilde{p}) = \bar{w} + \frac{r - g(\bar{w})}{g'(\bar{w})} \tag{A2}$$

and such that $\tilde{p} \gtreqless \bar{p} \Rightarrow \tilde{w} \gtreqless \bar{w}$

PROOF:

The function h is monotonic increasing. Differentiation of the right-hand side of (A2) shows that it is monotonic increasing in $\bar{w}$ provided $r>g$. This means that for any $\tilde{p}$ in the interval there exists a $\bar{w}$ which satisfies all the necessary and sufficient conditions for a maximum of the profit integral. It only remains to eliminate other alternatives. To prove: any optimal path must always be arbitrarily close to $\bar{w}$. Let w^* be an optimal path and define $\delta(\mathrm{t})=|w^*(\mathrm{t})-\bar{w}|$. From the necessary condition (9) we can then show that there exists an ϵ such that $|Dw^*(\mathrm{t})|=\epsilon(\mathrm{t})$ and ϵ is not arbitrarily small unless δ is. This is fairly clear intuitively from (9) but formal proof comes from a Taylor expansion of terms around $\bar{w}$. In addition, ***the sign of** Dw **is the sign of** $w(\mathrm{t})-\bar{w}$.* These two together mean that *if there is some τ such that $w^*(\tau)$ is not arbitrarily close to $\bar{w}$* then the optimal path $w^*(\mathrm{t})$ will move away from $\bar{w}$ for all $\mathrm{t}\geq\tau$. The rate of change $Dw^*(\mathrm{t})$ will not be arbitrarily small and will never become so because w^* is moving away from $\bar{w}$ for all $\mathrm{t}\geq\tau$. Hence w^* will reach w_{min} or w_M in finite time. But this cannot be optimal. Shutdown is not optimal because positive profits can be made. If the optimal path reaches w_M, the necessary condition says it can never get below w_M. But any path always above w_M violates the transversality condition and cannot be optimal. This provides a contradiction of the statement that $\delta(\mathrm{t})$ was not arbitrarily small for all t. Since any other optimal path cannot have any "daylight" between it and $\bar{w}$, then clearly $\bar{w}$ must be the unique optimal path.

PROPERTY 3(a). *If the price $p(\mathrm{t})$ is always above (below) $\tilde{p}$ then the optimal wage path $w^*(\mathrm{t})$ is always above (below) $\bar{w}$.*

PROOF:

This property is a straightforward extension of Property 2. We saw that the sign of $Dw(\mathrm{t})$ is the sign of $w(\mathrm{t})-\bar{w}$ when $p(\mathrm{t})=\tilde{p}$. This meant that if $w^*(\mathrm{t})$ ever went below $\bar{w}$ it would diverge to w_{min} and would not be optimal. This must hold a fortiori if $p(\mathrm{t})\geq\tilde{p}$ for all t. Similarly if $p(\mathrm{t})\leq\tilde{p}$ for all t, then w^* cannot go above $\bar{w}$ without diverging and again this is not optimal.

PROPERTY 3(b). *If $p(\mathrm{t})$ converges from above (below) to $\tilde{p}$, then the optimal wage path $w^*(\mathrm{t})$ converges from above (below) to $\bar{w}$, where $\bar{w}$ and $\tilde{p}$ are related by* (14) *and $p_{min}<\tilde{p}<p_{max}$.*

PROOF:

Define $u(\mathrm{t})$ and $v(\mathrm{t})$ by:

$$u(\mathrm{t}) = h[p(\mathrm{t})] - h(\tilde{p}) \tag{A3}$$

$$v(\mathrm{t}) = w(\mathrm{t}) + \frac{r - g(w)}{g'(w)} - h(\tilde{p}) \tag{A4}$$

Notice that from (12) $w=\bar{w}=>v=0$. The necessary condition (9) can be written:

$$\text{(A5)} \quad u(t) - v(t) = [g''(w)/g'(w)^2]Dw$$

Case 1: $p(t) \leq \bar{p}$. We know that $u(t) \leq 0$, $u(t) \rightarrow 0$ and, from (14) and Property 3(a), $v(t) \leq 0$. If there is ever a time τ such that:

$$\text{(A6)} \qquad |v(\tau)| - |u(\tau)| = \delta(\tau) > 0$$

then (A4) and (A5) show that this will be true for all $t \geq \tau$, and so $w(t)$ will decrease for all $t \geq \tau$. Such a path will reach w_{min} and cannot be optimal. Hence on the optimal path $|v(t)| \leq |u(t)|$ for all t. But since $u(t) \rightarrow 0$ as $t \rightarrow \infty \Rightarrow v(t) \rightarrow 0$ and $w^*(t) \rightarrow \bar{w}$.

Case 2: $p(t) \geq \bar{p}$ is proved in a similar way. It is slightly more complicated since if $w > w_M$ then $|v(t)|$ can become less than $|u(t)|$. It is still true that once $w > w_M$, it cannot get back, however, and this is all that is needed.

REFERENCES

K. J. Arrow and M. Kurz, *Public Investment, the Rate of Return, and Optimal Fiscal Policy,* Baltimore 1970.

M. N. Baily, "Wages and Employment under Uncertain Demand," *Rev. Econ. Stud.,* Jan. 1974, *41,* 37–50.

J. F. Burton, Jr. and J. E. Parker, "Interindustry Variations in Voluntary Labor Mobility," *Ind. Lab. Rel. Rev.,* Jan. 1969, *22,* 199–216.

J. P. Gould, "Adjustment Cost in the Theory of Investment of the Firm," *Rev. Econ. Stud.,* Jan. 1968, *35,* 47–55.

H. I. Grossman, "Aggregate Demand Job Search and Employment," *J. Polit. Econ.,* Nov./Dec. 1973, *81,* 1353–69.

J. J. McCall, "Economics of Information and Job Search," *Quart. J. Econ.,* Feb. 1970, *84,* 113–26.

D. T. Mortensen, "A Theory of Wage and Employment Dynamics," in E. S. Phelps et al., eds., *Microeconomic Foundations of Employment and Inflation Theory,* New York 1970, 167–211.

W. D. Nordhaus, "Recent Developments in Price Dynamics," in O. Eckstein, ed., *The Econometrics of Price Determination,* Washington 1970, 16–49.

E. S. Phelps and S. G. Winter, "Optimal Price Policy under Competition," in Phelps et al., eds., *Microeconomic Foundations of Employment and Inflation Theory,* New York 1970, 309–37.

E. S. Phelps et al., *Microeconomic Foundations of Employment and Inflation Theory,* New York 1970.

L. S. Pontryagin et al., *The Mathematical Theory of Optimal Processes,* New York 1962.

P. A. Samuelson, *Economics,* 7th ed., New York 1967.

S. O. Schweitzer, "Factors Determining the Interindustry Structure of Wages," *Ind. Lab. Rel. Rev.,* Jan. 1969, *22,* 217–25.

G. J. Stigler, "Information in the Labor Market," *J. Polit. Econ.,* Oct. 1962, *70,* 94–105.

L. Ulman, "Labor Mobility and the Industrial Wage Structure of the Postwar United States," *Quart. J. Econ.,* Feb. 1965, *79,* 81–85.

Organization for Economic Cooperation and Development, OECD, *Wages and Labor Mobility,* Paris 1965.

Academic Achievement and Job Performance

By David A. Wise*

This paper examines the relation of measures of academic achievement and other personal characteristics to job productivity of college graduates working in a large corporation. It was motivated by a desire to study the effectiveness of resource use in institutions of higher education.

Numerous studies of returns to investment in human capital have demonstrated quite clearly that the more educated earn more than those with less education.[1] These studies, however, do not tell us why this is so. Edgar Friedenberg has posed the issue as follows:

> Do schools contribute to economic opportunity? It can certainly be demonstrated that the longer one stays in school, up to 20 years at least, the higher one's income is likely to be. But whether this is due to capacities actually developed through instruction, or the effect on character and personality of decades of submission to school routines, or merely the consequence of a complex, interlocking series of credentials which restrict opportunities to those who have satisfied the authorities at an earlier stage is not clear. All these processes are involved and are interrelated. But I have listed them, I think, in ascending order of their influence, though apologists for the educational system would prefer, I believe, that the order be reversed. [p. 98]

Entry into higher level jobs is often restricted to college graduates, and in many cases graduate or professional degrees are required. Persons are selected and certified in the higher educational system largely on the basis of measures of academic aptitude or performance; but the relation between these measures and job performance (or productivity) is not generally known. The assumption of economic efficiency, however, would imply the existence of a causal relationship and equity would require it. A requisite of equity, or fairness, its intended meaning, would presumably be that persons with the same productive potential in an occupational field have the same opportunity to enter it. Such a requirement would not be met if persons were selected for jobs on the basis of "nonproductive" traits. This gives rise to two questions: First, is there a relation between an individual's ranking as determined by measures used to select and certify within institutions of higher education and his performance on jobs outside the educational system? Second, if there is a relation, why is it observed?

It may be useful to relate these questions to the notion of credentialism. A credential required by an employer could be, for example, a degree, a minimum class rank or grade point average, graduation from a college of at least some minimum quality, or a combination of these. We would like to know whether or not a college degree represents more than simply a credential

---

* Assistant professor, Harvard University. This research was supported by the Ford Foundation Program for Research in University Administration. I have benefited from many discussions with Roy Radner, for whose advice and suggestions I am especially grateful. I would also like to thank Lloyd Ulman, Paul Taubman, George Borts, and an anonymous referee for their very helpful comments. Responsibility for remaining shortcomings, of course, remains with me. In addition, I acknowledge the generous and invaluable assistance of Paul Banas, Personnel Research Section, Fort Motor Company, in obtaining the data used in the study.

[1] See, for example, Gary Becker, Giora Hanoch, Lester Thurow (1967).

requirement. Is it also an indicator of productive ability? If it is, then we might expect degrees of "different quality," as measured by differences in academic achievement, to indicate different levels of productive ability. If the degree is an indicator of productive ability, we would like to know, in addition, whether or not a college education contributes to one's productive ability. This is the essence of the second question.

I. Background and Outline

A. *The Initial Question*

Selection in the higher educational system is currently based primarily on measures of cognitive abilities[2]—grades,[3] achievement tests, and aptitude tests.[4] In short, students tend to be selected on the basis of their promise as future students in the existing educational structure. Certification involves primarily the assigning of grades (presumably to indicate the degree of mastery of academic subject matter) and the awarding of degrees along with verification that a certain amount of time has been spent in school and specific courses.

Several authors have concluded, however, that these measures are not correlated with job performance in many occupations, even though they may be used in occupational selection.[5] The initial and primary focus of this paper is an examination of this conclusion when applied to a sample of individuals who work in a large corporation, not unlike many others. The absence of a significant relation between academic achievement and job performance may be taken as a null hypothesis. This hypothesis is rejected in this case, and we are led to ask why.

B. *The Second Question*

According to human capital theory,[6] an individual chooses the occupation and level of education that maximize the present value of his expected lifetime earnings. It is generally assumed that education changes an individual in such a way as to increase his capacity to perform job-related tasks. Although not necessary to the theory, it is also implicitly assumed that the knowledge and other cognitive skills gained in school increase his productive capacity.[7] But it may be that this is not the primary reason for observed differences in earnings by educational level. The educational system may select persons who are more productive, but not increase their productive potential; it may serve simply as a screening device. Recent theoretical examples of this have been presented by Michael Spence and Kenneth Arrow. And if education does increase productive potential, it may be by way of changes in affective rather than cognitive attributes. This possibility is not inconsistent with human capital theory, but may suggest changes in emphasis within the educational system.

---

[2] The terms "cognitive," "affective," "academic," and "nonacademic" are used rather loosely throughout the paper. Cognitive and affective are used to distinguish personal traits which Benjamin Bloom describes in terms of educational objectives as follows:

> The cognitive domain . . . includes the objectives which deal with the recall or recognition of knowledge and the development of intellectual abilities and skills. . . .
>
> The affective domain . . . includes objectives which describe changes in interest, attitudes, and values, and the development of appreciations and adequate adjustment. [p. 7]

The terms academic and nonacademic are often used in referring to cognitive and affective traits, respectively. The distinctions are quite imprecise and should be interpreted accordingly.

[3] Grades may not measure only cognitive abilities. See Herbert Gintis, for example.

[4] See *The Student in Higher Education*. See also the *Report of the Commission on Tests* of the *CEEB*, Leonard Baird and John Holland, and Baird and James Richards.

[5] In his study relating primarily (but not exclusively) to blue collar workers, Ivar Berg claims that there is little or no relation between job performance and education. Another investigator, Donald Hoyt, reviewed 46 studies relating college grades and adult achievement in several areas and in the summary stated ". . . present evidence strongly suggests that college grades bear little or no relationship to any measures of adult accomplishment." Similar conclusions have been suggested by others. See, for example: Seth Payne, who remarks on studies by the National Advisory Committee for Aeronautics (now part of *NASA*) and Hughes Aircraft Company; Calvin Taylor and Robert Ellison; Donald MacKinnon; D. A. Goslin (1963, 1968); Dael Wolfle.

[6] See Becker or Thurow (1970). The conventional theory is summarized by Yoram Weiss in a more recent article.

[7] In fact it may be explicitly stated. Becker, for example, states: "On-the-job and school training are not the only activities that raise real income primarily by increasing the knowledge at a person's command" (p. 31). See also p. 86 of the same source.

The screening hypothesis suggests the possibility that higher earnings of the more educated result in part from credential requirements which restrict entry to many occupations on the basis of educational level. Even in an extreme case in which (college) education does not change the productive capacity of students in any way, a positive relation between earnings and education might be observed. In general, what is required is that the cost of education to an individual be a decreasing function—or that the availability be an increasing function—of productive traits, that employers set wages to maximize expected profits, and that individuals choose an educational level to maximize expected earnings.[8]

Several investigators[9] have attempted to control for intelligence in studies relating earnings to education. In general, these studies suggest a rather low contribution of intelligence (as measured) to earnings when level of education is controlled for.[10] This may, however, be due in part to occupational entry requirements based on level of education and limits on potential earnings within occupations. In addition, earnings may not be highly related to performance in occupations in which earnings are largely determined by established wage scales and promotions are based on seniority.

Gintis has argued that noncognitive traits which lead to academic achievement are also required for success in bureaucratic organizations. He contends that it is these affective traits fostered by the educational experience (or at least rewarded by the educational system), and not cognitive development, which are largely responsible for the increased earning power of the more educated. His argument also neglects credential requirements, which may limit the explanatory power of cognitive ability when educational level is controlled for.

After examination of the relation of measures of academic achievement and other personal characteristics to job productivity in Section II, subsequent discussion in Section III will explore implications of the findings with respect to these questions.

C. *The Data*

This investigation is based on the performance of individuals with a given level of education, but different levels of academic achievement, who work in a particular environment and perform similar tasks. The data pertain to individuals working in a large manufacturing corporation employing both technical and nontechnical personnel. The sample included persons employed in all parts of the country, although a large proportion were employed in a single region. In 1968 the corporation obtained biographical information and data on salary and position in the corporation of approximately 1300 college graduates employed at that time. All were hired before 1965, and all were not more than 30 years old when they were hired. Thus, most had little or no previous job experience. Those surveyed represented a stratified random sample (by function—manufacturing, finance, sales, etc.—and hierarchical position in the firm) of approximately 6800 persons who met these requirements. All persons in the sample were white males. Biographical information pertained to socioeconomic background, high school and college nonacademic activities, academic performance (grades), college attended, employment goals, early employment experiences with the firm, and other personal characteristics. Data on salary, position in the firm, years employed, and college attended were obtained from personnel records; all other information was obtained by survey. Excluded from the sample were: persons hired before 1945, because the relation between salary and years employed differed markedly between this group and the group hired later; and persons who had an M.A. degree when they joined the firm.

---

[8] Spence presents several examples in which the cost of education is related to a productive trait. Similar examples are presented by Arrow.

[9] See, for example, Zvi Griliches and William Mason, Otis Duncan, Carl Bajema, Orley Ashenfelter and Joseph Mooney, and Lee Hansen, Burton Weisbrod, and William Scanlon.

[10] The study by Hansen, Weisbrod, and Scanlon is an exception, as is the more recent work of Paul Taubman and Terence Wales.

D. *The Measurement of Job Performance*

There seems to be no completely satisfactory way of measuring job performance, or even of defining it. We therefore resort to the assumption that differences in individual job performance are reflected in measures of success within the firm; that rewards within the firm are based on job performance, or at least on perceived performance. In this case, there are two available measures of success: salary and grade level. The model of individual experience within the firm proposed below is intended to be consistent with salary as a measure of success. If in fact, as is commonly assumed, an individual's lifetime earnings reflect his marginal product, the link between salary and job performance is rather direct. However, the relation between individual attributes and reward by the firm is of interest whether or not earnings reflect marginal product.[11]

E. *Summary of Results*

The relation between college quality and grades on the one hand and this measure of job performance on the other is not only statistically significant but is quantitatively important. Thus, it would appear that criteria used for selection and certification are positively associated with an individual's ability to perform job-related tasks. The findings of the study suggest that this relation is not simply due to noncognitive attributes such as motivation or to "innate ability" such as IQ, which may underlie academic success; but that academic achievement is an important determinant of job performance. The existence of such a relationship is not, however, directly demonstrated here; only indirect evidence is provided. A direct measure of IQ, for example, is not available.

The results also suggest that nonacademic attributes are as important as academic abilities in determining job performance. Leadership ability and initiative (or job experiences that allow imaginative thinking or expression of one's ideas) are found to be positively associated with job performance, while a negative relation is observed between an individual's desire for job security and his job success. An important characteristic of the sample data was the low correlation between academic and nonacademic variables, so that identification of separate effects of these variables was not hampered.

---

[11] Another model, which is used with grade level as a measure of success, has been suggested elsewhere by the author (1972, 1975), but is not discussed in detail here. It is based on the assumption that persons move through grade levels within the firm or leave the firm according to a Markov scheme and that relative job performance is reflected in the probability of promotion from one grade level to the next.

II. Salary and Personal Attributes

A. *The Variables*

Job productivity is assumed to be influenced by a variety of cognitive and affective traits, none of which is measured precisely by available data. Those variables assumed to reflect primarily academic aptitude or achievement are high school grade point average, college selectivity, college grade point average, and rank in graduate class for persons who obtained an M.A. degree after joining the firm. The variables assumed to measure nonacademic attributes are indices that were obtained by combining responses to groups of biographical questions. The indices are referred to as: socioeconomic index, leadership and organizational ability index, initial experience (or initiative index), initial supervisor experience, and job security index. Each index represents the first principal component relative to responses to a group of questions. For convenience, each index has been scaled to have mean zero. This procedure and the rationale for its use have been discussed by the author in some detail elsewhere.

It is assumed that socioeconomic background affects not only the quality of education an individual obtains, but it may also have some independent effect on nonacademic traits which influence later job performance in a corporate (or bureaucratic) environment.[12] The leadership and organization ability index is assumed to characterize, in part, an individual's past demonstrated ability as a leader or organizer.

The initial experience index may be interpreted in two ways. It may indicate the existence of a particular environment during an individual's first years with the firm. But it may be more likely to reflect an individual's personal behavior; in this case traits such as initiative, decision-making ability, or imaginative thinking. The initial supervisor index seems to be less indicative of the respondent's behavior and more representative of his initial work environment—the extent to which his initial supervisor included him in the decision-making process and pushed for his advancement.

The job security index (based on the response to only one question) may be interpreted in two ways. First, as a measure of risk aversion. The higher a person progresses in the firm hierarchy, the greater is the likelihood that he will have to make decisions, the consequences of which are traceable to him.[13] A second possibility is that the response simply reflects an individual's confidence or "self-assurance" (see Ghiselli). Persons who are not worried about job security may feel confident that they can make the right decisions.

The correlations between academic and nonacademic variables (excluding the socioeconomic index) are quite small in most cases.[14] The two groups of variables seem to reflect largely independent individual (or job environment) attributes.

---

[12] See, for example, Edwin Ghiselli who emphasizes the importance of "perceived occupational level" as a determinant of managerial success. This might be influenced by the socioeconomic background of one's parents. See also Edwin Robinson, Robert Athanasiou, and Kendra Head, and M. Winterbottom.

[13] In this regard, a firm employee stated that at particular grade levels, persons tend to come under "scrutiny" by higher management.

[14] The correlations are not shown here, but interested readers may obtain them from the author. The low correlations between academic and nonacademic variables are consistent with the findings of other investigators. See, for example, Jacob Getzells and Philip Jackson and MacKinnon, who find weak relations between intelligence and creativity; Holland and Richards (1965, 1966), and Holland and R. Nichols, who find that academic and nonacademic (extracurricular) achievement in high school and college are "relatively" independent. Similar evidence is provided by Baird and Holland, and J. Flanagan et al.

B. *The Model*

In what follows it is implicitly assumed that differences among individuals in expected lifetime earnings reflect differences in the performance of tasks considered important to the functioning of the firm. Those who "do better" are rewarded with higher incomes. A more conventional assumption would be that the marginal product of a person who stays in the firm is given by his expected lifetime earnings. The results below, however, are in no way dependent on an individual being paid the value of his marginal product.

Let us assume that the monthly salary of a person who has been employed by the firm for t years is given by

$$s = e^a e^{rt} e^{\epsilon} \tag{1}$$

where e^a represents starting salary, r is the rate of increase in monthly salary, and ϵ is a disturbance term. (As indicated below, this functional form is used because it is more consistent with the experience of persons in the sample than are other alternatives.) Both a and r are assumed to be functions of personal characteristics:

$$a = g_0(x) \tag{2}$$

$$r = g_1(x) \tag{3}$$

where x is a vector of individual (and possibly job environment) attributes. Not all elements of x will enter both g_0 and g_1 with nonzero coefficients.

The estimated model allows starting salary and rate of salary increase to vary with personal characteristics according to particular functional forms (see (5) and (6) below). Taking the logarithm of (1) and substituting these functional forms, we obtain the estimated model (7) as follows:

$$ln\ s = a + rt + \epsilon \tag{4}$$

$$a = g_0(x) = a_0 + a_i + b_j + dx_0 \tag{5}$$

$$r = g_1(x) = r_0 + \alpha_i + \beta_j + \gamma_k + \Sigma_{l=1}^{5} \delta_l x_l \tag{6}$$

where the following identifying restrictions are imposed.

$$\Sigma_i a_i = 0, \qquad \Sigma_j b_j = 0, \qquad \Sigma_i \alpha_i = 0,$$
$$\Sigma_j \beta_j = 0, \qquad \Sigma_k \gamma_k = 0$$

Then,

$$ln\ s = a_0 + a_i + b_j + dx_0 + [r_0 + \alpha_i + \beta_j + \gamma_k + \Sigma_{l=1}^{5} \delta_l x_l]\text{t} + \epsilon \tag{7}$$

The analysis of covariance formulation depicted in (7) was estimated by ordinary least squares.

As above, s represents monthly salary and t, years employed; d and δ are coefficients. The other symbols are defined in the list below. All but x are parameters to be estimated. The mnemonic abbreviations to the left of the symbols are used throughout in tables and tabulations.

Constant $= a_0 \equiv$ constant
BA $= a_1 \equiv$ effect of having B.A. degree when start work at the firm
BA' $= a_2 \equiv$ effect of not having B.A. degree when start work at the firm
ENG $= b_1 \equiv$ effect of engineering or science undergraduate major
LIB $= b_2 \equiv$ effect of liberal arts (or other) undergraduate major
BUS $= b_3 \equiv$ effect of business undergraduate major
EXP $= x_0 \equiv$ years of experience (excluding military) before coming to the firm
AV $= r_0 \equiv$ "average" rate of salary increase
SEL(i) $= \alpha_i \equiv$ effect of undergraduate college in *i*th selectivity group

$GPA(j) = \beta_j \equiv$ effect of undergraduate grades in jth interval

$MA(k) = \gamma_k \equiv$ effect of being in kth rank in M.A. class (γ_0 is for B.A. only)

$SEC = x_1 \equiv$ index of desire for job security

$SES = x_2 \equiv$ socioeconomic background index

$LEAD = x_3 \equiv$ leadership and organizational ability index

$INT = x_4 \equiv$ initial job experience index

$SUP = x_5 \equiv$ initial supervisor index

A few persons joined the firm before they received a bachelor's degree; they are distinguished from those who had a degree when hired. Information on job experience before joining the firm was not available. The previous experience variable is a crude estimate made from available information. Colleges were assigned to groups on the basis of Alexander Astin's (1965b) college selectivity (*SEL*) index.[15] This index is intended as a "measure of the average ability level of the entering class." The grade point average (*GPA*) intervals are those designated in the firm survey.[16] Rank in graduate school pertains to those who obtained a master's degree after joining the firm. Those in group 1 graduated in the top 5 percent of their graduate class; group 2, in the top one-third; and group 3, in the lower two-thirds. The final group, indexed by 0, represents those with a B.A. degree only. The last five variables, which have been discussed earlier, are treated as continuous.

It may be noted that the model as stated does not allow selectivity or college grades to enter into the determination of starting salary; they only affect the rate of salary increase. This is not an arbitrary assumption but is based on previous analysis of the data (see my earlier paper). Appropriate formulations of the model indicated that starting salary was not appreciably affected by college selectivity or grade point average, contrary to a priori expectations. Firm employees confirmed that, in the interest of morale, these variables were not normally considered in determining starting salary. Of course, it could be that the firm did not believe that differences in college selectivity and grade point average among persons hired represented significant differences in productive ability. Models allowing prior military service to influence starting salary and undergraduate area of study to affect the rate of salary increase were also tried. These factors seemed unimportant and were eliminated from later analysis.

In addition, the model as stated does not allow for interactions between college selectivity and college grades (or between rank in graduate school and either of these). This was not feasible with six selectivity groups and four grade point average groups; but the significance of interactions was tested for in a model with only three groups for each. Interactions were shown not to be statistically significant in this case, and it has been implicitly assumed that they would not be significant if they were included in the model with more groups.

Finally, the model assumes a constant rate of salary increase over an individual's working lifetime. It is normally assumed that the rate of increase decreases and may become negative at some point. The data used here did not exhibit this characteristic, although a model with salary, rather than the logarithm of salary, expressed as a linear function of years employed yields results comparable to those obtained with the above model. It should be remembered that the available data pertain to persons employed no longer than twenty-two years.

---

[15] The groups and corresponding intervals are as follows: *SEL1*, 73–79; *SEL2*, 66–72; *SEL3*, 61–65; *SEL4*, 56–60; *SEL5*, 47–55; *SEL6*, 37–46.

[16] The groups and corresponding intervals are: *GPA1*, 3.50–4.00; *GPA2*, 3.00–3.49; *GPA3* 2.50–2.99; *GPA4* less than 2.50. *GPA4* combines the lowest two intervals as designated in the survey.

It should also be kept in mind that data are available on salary only, which may comprise only a portion of total income from the firm. As one rises in the firm hierarchy, bonuses and stock options are likely to comprise an increasingly large portion of total income and may even be more important than salary at very high levels.[17]

C. *Results*

Estimates of the parameters in equation (7) obtained by ordinary least squares[18] are shown in Table 1. The F-statistics pertain to the null hypothesis that the parameters indicated are all zero, which in the case of the discontinuous variables is equivalent to the hypothesis that the effects are the same for all groups (for example, all college selectivity groups). In each case, the hypothesis is rejected at any reasonable level of significance.[19] Estimates given in the upper portion of the table relate to determinants of starting salary a, and those below to the rate of increase in salary r.

The estimates regarding starting salary confirm what might be expected. Persons with a bachelor's degree had a higher initial salary than those who started work before obtaining a degree. The estimates also indicate that persons whose undergraduate major was engineering or a science commanded higher initial salaries than those who majored in liberal arts or business. A comparison of the previous experience coefficient with the average rate of salary increase indicates the relative value the firm places on experience prior to joining the firm versus experience in the firm.[20]

TABLE 1—PARAMETER ESTIMATES, TOTAL SAMPLE

Variable	Estimated Coefficient	Standard Error[a]	F-Statistic
Constant	6.70356		
BA	.02591	(.01079)	$F=5.766$
BA'	−.02591	(.01079)	
ENG	.03879	(.00764)	$F=16.018$
LIB	−.01340	(.00956)	
BUS	−.02539		
EXP	.01647	(.00315)	
AV	.04501	(.00171)	
SEL1	.01085	(.00308)	$F=10.781$
SEL2	.00234	(.00119)	
SEL3	.00122	(.00147)	
SEL4	−.00183	(.00091)	
SEL5	−.00431	(.00116)	
SEL6	−.00827		
GPA1	.00777	(.00131)	$F=22.179$
GPA2	.00055	(.00087)	
GPA3	−.00245	(.00073)	
GPA4	−.00587		
MA1	.01241	(.00311)	$F=11.398$
MA2	−.00017	(.00182)	
MA3	−.00504	(.00193)	
MA0	−.00720		
SEC	−.00310	(.00045)	$F=41.893$
SES	.00030	(.00031)	
LEAD	.00218	(.00027)	
INT	.00070	(.00025)	
SUP	.00142	(.00035)	
$N=976$			
$R^2=.69029$			

[a] The standard error for the estimate of the effect of being in the lowest interval of a variable (e.g., *SEL6*) is not shown because the computer program used did not compute a variance-covariance matrix.

---

[17] A plausible functional relation between total earnings (E) and salary (s) is $E=s^k$, where k is some constant greater than one. In this case, the use of total earnings instead of salary would not qualitatively affect the estimates of the parameters, a and r in (1). They would simply be increased by a factor of k. Differences in total lifetime earnings among individuals would, of course, be much larger than differences in lifetime earnings from salary.

[18] One would expect heteroscedasticity in a regression of salary on the right-hand variables included here. Using $\ln s$ instead of s tends to reduce the problem, but does not eliminate it, judging from a plot of residuals against years employed. Attempts to correct for this, however, did not appreciably alter the parameter estimates or their standard errors.

[19] $F_{.01}(5, 976)=3.02$, $F_{.005}(5, 976)=3.35$.

[20] According to these estimates, the average person with x years of experience outside the firm would receive an initial salary of $ke^{.016x}$ (where k is a constant)

(over)

TABLE 2—ESTIMATED RATES OF SALARY INCREASE FOR PERSONS WITH A B.A. DEGREE BY COLLEGE SELECTIVITY-COLLEGE GRADE-POINT AVERAGE GROUP

Without *SEL-GPA* Interactions[a]

	SEL1	*SEL2*	*SEL3*	*SEL4*	*SEL5*	*SEL6*
GPA1	.05643	.04792	.04680	.04375	.04127	.03731
GPA2	.04921	.04070	.03958	.03653	.03405	.03009
GPA3	.04621	.03770	.03658	.03353	.03105	.02709
GPA4	.04279	.03428	.03316	.03011	.02763	.02367

With *SEL-GPA* Interactions[b]

	SEL1-2	*SEL3-4*	*SEL5-6*
GPA1-2	.04395	.03806	.03442
GPA3	.03794	.03423	.02923
GPA4	.03448	.03078	.02607

[a] Calculated from estimates in Table 1, $[AV+MA_0+SEL(i)+GPA(j)]$.

[b] Calculated from a regression equation allowing for *SEL-GPA* interactions.

The "average" rate of increase in salary is estimated to be .045. The estimated rate of salary increase goes up consistently with college selectivity, college grades, and rank in graduate class (if a higher degree is obtained). The relationships, however, are seen to be non-linear.[21]

The effect of graduate education after joining the firm can be seen by combining the average rate of increase with the effects for rank in graduate school. The average rates of increase by rank in graduate class are as follows:

Rank in Graduate Class	Estimated Rate of Salary Increase
Top 5 percent	.057
Top ⅓, excluding top 5 percent	.045
Bottom ⅔	.040
B.A. only	.038

According to these estimates, obtaining a master's degree has almost no effect on an individual's rate of salary increase, unless he graduates at least in the top third of his class. Simply obtaining a certificate has no appreciable effect.

From the parameter estimates in Table 1, the implied rates of salary increase by selectivity and grade point average group for persons with a bachelor's degree only have been calculated. They are shown in the first half of Table 2. The estimate in the *i*th column and the *j*th row is obtained as follows: $AV+MA_0+SEL(i)+GPA(j)$. The estimated rate for persons from the highest selectivity schools and with the highest grades is more than twice as high as the rate of increase for persons from the lowest selectivity schools and with the lowest grades. These are not independent estimates. For comparison, independent estimates for fewer categories calculated from parameter estimates from a regression with selectivity and grade point average interactions are shown in the second half of Table 2.

The estimates of the parameters corresponding to the nonacademic variables are all significant except the coefficient on socioeconomic background. Apparently,

---

while a similar person with z years of experience within the firm and no previous experience would be earning an amount $ke^{.045z}$. Although the coefficient on *EXP* may be underestimated, it might be expected to be quite low, because the previous experience for many people in the sample may have been in work unrelated to that of the firm. In addition, part or all of the estimated previous "experience" for an individual could have been time spent without a job (after college) or time spent looking for work. In any case, the estimated coefficient should be interpreted cautiously.

[21] The apparent nonlinearity in selectivity may result in part from the way in which the index was standardized (see Astin (1965b)).

its influence is only by way of its effect on academic achievement. The coefficient is significant when college selectivity is left out of the equation.

One way to evaluate the relative influence of the continuous variables is to compare changes in the rate of salary increase Δr, resulting from standard deviation increases. These changes are shown in the following tabulation.

Variable	Sample Standard Deviation	Δr
SEC	1.130	−.00350
SES	1.452	.00044
LEAD	1.557	.00339
INT	2.113	.00148
SUP	.1487	.00211

The change due to a standard deviation increase in the job security or leadership index is approximately the same as the increase resulting from a shift from the fourth to the third, or from the third to the second, college grade groups. The effects of the initial experience (or initiative) and initial supervisor indices are considerably weaker. The latter variables presumably pertain only to an individual's initial experience in the firm, whose effects may tend to disappear as the number of years employed increases.[22]

An idea of the relative contribution of academic versus nonacademic variables in the "determination" of salary may be obtained by comparing values of R^2 obtained when different groups of variables are included in the regression. Selected groups of variables and corresponding R^2 values are as follows.

Right-Hand Variables Included	R^2
1) All Variables	.690
2) t (years employed) only	.490
3) t, plus variables affecting starting salary: *BA*, *ENG*, *LIB*, *EXP*	.530
4) t, *BA*, *ENG*, *LIB*, *EXP*, plus academic variables: *SEL*(*i*), *GPA*(*j*), *MA*(*k*)	.622
5) t, *BA*, *ENG*, *LIB*, *EXP*, plus nonacademic variables: *SEC*, *SES*, *LEAD*, *INT*, *SUP*	.633

The percentage of the variation remaining unexplained after variables 3) are included, which is explained by the academic and nonacademic variables together, is .340. $[(.690-.530)\div(1-.530)=.340]$. A high estimate of the contribution of academic variables is the proportion of remaining variance explained when they are added to the regression in the absence of nonacademic variables; a low estimate is the addition to the proportion explained when they are added after the nonacademic variables.[23] In this case, the high and low estimates are $[(4)-(3)]/[1-(3)]$ and $[(1)-(5)]/[1-(3)]$, respectively. The opposite is true for the nonacademic variables. This yields the following results: academic variables—low, .121; high, .196 and nonacademic variables—low, .145; high, .219.

D. *Undergraduate versus Graduate School Achievement*

Further insight into the relation between salary progress and graduate education may be gained by considering the mean of a salary index (described in the note to Table 3) by undergraduate grade point average (*GPA*) and rank in graduate school (*MA*). These figures are presented in Table 3.[24] They confirm the findings

---

[22] The appropriate interpretation of these variables is also less clear than that of *LEAD* or *SEC*. If, for example, *INT* is assumed to measure initiative, it may do so with considerable error, thus tending to "bias" the estimated coefficient toward zero. In other words, the extent of the error in variables with respect to *INT* and *SUP* is likely to be larger than for *SEC* and *LEAD*.

[23] The two estimates would be the same, of course, if the two groups of variables were orthogonal. In this case, the most highly correlated variables are *SES* and *SEL* (−.22), *SEC* and *SEL* (.20), and *SUP* and *MA* (*k*) (.23).

[24] Other variables have not been controlled in making these estimates. It might be assumed, for example, that

TABLE 3—MEAN OF SALARY INDEX BY UNDERGRADUATE MAJOR AND *GPA-MA* GROUP FOR PERSONS WHO OBTAINED AN M.A. DEGREE[a]

	Total Sample		Engineering or Science		Business	
	MA1–2	*MA3*	*MA1–2*	*MA3*	*MA1–2*	*MA3*
GPA1–2	23.26 (4.12) $n=39$	10.85 (4.12) $n=15$	34.97 (7.75) $n=13$	18.78 (10.55) $n=7$	16.04 (4.90) $n=21$	2.23 (10.04) $n=5$
GPA3–4	11.90 (4.29) $n=36$	3.44 (4.48) $n=33$	20.91 (7.21) $n=15$	0.50 (7.21) $n=15$	5.29 (5.61) $n=16$	1.91 (6.23) $n=13$
Total:		13.10 (2.39) $n=123$		18.14 (4.25) $n=50$		8.31 (3.06) $n=55$
R^2		.084 $F(3, 119)=3.65^*$		.190 $F(3, 46)=3.62^*$		.076 $F(3, 51)=1.40$

Model: $e_{ijk}=\beta_{ij}+\epsilon_{ijk}$, where e_{ijk} is the index number for the kth person in the ijth *GPA-MA* group and β_{ij} is the mean of the index numbers for persons in the ijth *GPA-MA* group.

[a] The index is the residual from a regression of salary on years employed where a correction has been made for heteroscedasticity. The total sample was used in this regression. The index would of course have mean zero over all individuals in the sample. As can be seen, M.A. holders have salary index numbers greater than zero on the average in all *MA-GPA* groups. Figures for liberal arts majors are not shown because of the small number of observations in each cell. Numbers in parentheses are standard errors of estimates. (They are to be interpreted the same as standard errors in any regression model; in this case β_{ij} are parameters estimated.) For example, the mean of the salary index for engineers in group *GPA1–2/MA1–2* is 34.97. The standard error of this estimate is 7.75. The F-statistic relative to the hypothesis that the means of all four engineering groups are the same is 3.62. A * indicates that the hypothesis is rejected at the .05 level of significance. The number of observations in the group is n.

presented in Table 1, but suggest a considerably larger effect of the master's degree on earnings for engineers or scientists than for students of business administration. In this case, one might expect a stronger effect for engineers because of the relatively unique combination of technical and managerial skills which they might be expected to possess.

III. Further Implications

The above analysis has been directed primarily toward the following question. What is the relation, if any, between academic achievement and other nonacademic attributes of college graduates and job performance (or productivity) outside the educational system? A consistent positive relation between commonly used measures of academic achievement and rates of salary increase has been found. Of those attributes of individuals that have been controlled for, academic and nonacademic characteristics seem to be of approximately equal importance in explaining differences in salary. Both, however, explain a relatively small proportion

---

persons with better undergraduate records were able to get into better graduate schools and that the difference in salary by undergraduate *GPA* is partially the result of this effect. Since, however, these persons obtained an M.A. at the same time that they were working, it is likely that the best graduate schools were not represented in the sample, and that the difference in the quality of those represented was not large.

of the total variation in salary. This is not surprising, since the individuals all had the same quantity of education when they started work and all were presumably selected with the expectation that they would be able to perform at least adequately. (If persons with poor academic credentials were hired because they were strong on other attributes not controlled for, the observed effect of academic characteristics would tend to be biased downward.) The Markov model explored elsewhere by the author (1972, 1975) explicitly implies a limit on the explanatory power of individual attributes.

Observed differences in salary are of course affected by on-the-job training, as well as academic training and other personal attributes. (See Jacob Mincer, for example.) This type of learning has not been accounted for here.

In addition to the above findings, the results may also provide some insight into the more difficult and more fundamental questions that were posed at the beginning of the paper. The implications with respect to these questions are less clear than the implied relation between academic achievement and job productivity. They essentially ask why this relation is observed.

A. *Intelligence or Academic Aptitude*

Let us consider first the possibility that success in mastering academic subject matter is not a determinant of job performance. Rather the observed relation between college grades and salary results from the fact that persons who got good grades had higher innate intelligence or academic aptitude than those who did not.[25]

We do not have a good control for academic aptitude. However, the variation in academic aptitude (Scholastic Aptitude Test (*SAT*) scores) among students in a college decreases with it's selectivity. There is little variation in the most selective schools where all students have high aptitudes. But even among persons from the best schools we observe a consistent effect of college grades. Looking at the second half of Table 2 we observe that the effect of grades on the rate of salary increase seems not to depend on selectivity. The difference between the estimated effects of the highest and lowest college grade groups is .0095 in the highest selectivity group and .0084 in the lowest. (The numbers represent differences in rates of salary increase.) If grades were serving only as a proxy for aptitude, one might expect the observed effect of grades to decrease with college selectivity.[26] "Lifetime earnings" not only increase with the rate of increase in salary, but do so at an increasing rate.[27] The higher the selectivity group, the higher the rate of increase in salary for any grade point average level. It could be that differences in grades represent relatively larger differences in knowledge in better schools. It is also possible that "smarter" persons can make better use of knowledge on the job.

---

[25] It is, of course, unreasonable to think that intelligence alone determines job performance. Most jobs require mastery of certain skills such as reading and writing. But we are here considering the effects of college education, and are referring to intelligence given that a person obtained a high school degree. Even in this case, some skills (e.g., techniques for engineers) may be necessary; but the mastery of these skills above some minimal level may not be relevant.

[26] Average *SAT* scores at the best schools are very high and there is very little variation among the scores. See, for example, Astin (1971). It should also be noted, however, that differences in the variation of *SAT* scores may not be an accurate representation of differences in the variation of underlying aptitude. This may be especially true at better schools where many persons score at the top of the *SAT* scale. In addition, these scores represent an ordinal ranking and not a cardinal measure.

[27] This is a property of the functional form used, and should not be overemphasized. To demonstrate it, let T be the number of years a person works (we are limited by the data to 22 years), k be starting salary, and let LE represent expected lifetime earnings (i.e., the integral from 0 to T of ke^{rt}). It may be seen that the second derivative of lifetime earnings with respect to r must be greater than zero.

It has been implicitly assumed that the observed relation between rate of salary increase and college selectivity is primarily due to the higher average scholastic aptitude of persons from better schools. It is also possible that persons from better schools learn more than those in poorer schools, so their accrued academic knowledge is greater.

B. *Affective Traits*

It has been suggested that observed differences in job performance by grade point level may be due to nonacademic traits which are possessed to a greater extent by persons who do well in school than by those who do poorly (see Gintis, for example). An attempt has been made to control for some of these traits in the estimation procedure by incorporating the indices for security, leadership, socioeconomic background, and initiative—certainly an incomplete representation. However, it seems reasonable that job productive traits not controlled for, which also influence performance in college, would affect high school performance.[18] If college grades are serving only as a proxy for excluded affective traits, high school grades may be a reasonable proxy as well. Then if high school grade point average were introduced into the regression equation, the effects of college and high school grades should be confounded, at least to some degree. When this is done, however, we find that high school grades are not significant and the estimated effects for college grades are affected only slightly. (Interested readers may obtain these results from the author.) At least some reduction might be expected since high school and college grades may both be proxies for academic aptitude (not completely controlled for by college selectivity) and may also be substitute measures for some kinds of knowledge.

It might be expected that traits associated with achievement in undergraduate school and an M.A. program are even more closely related than traits associated with high school and undergraduate achievement. But Tables 1 and 3 suggest that the extent to which an individual masters graduate subjects affects his productivity, even if he has demonstrated the possession of traits associated with high grades, by getting high grades as an undergraduate, and vice versa. This does not support the hypothesis that college grades are serving simply as a proxy for noncognitive characteristics.

In contrast, assume that the relation between education and job performance is due to the acquisition of knowledge (of academic subject matter) during the educational process. It might be expected that given a measure of an individual's mastery of course work in college, an additional measure of how well he did in high school would not add much to an overall measure of his knowledge of academic subject matter. Knowledge acquired in college is likely to dominate that acquired in high school. Under this assumption, given college grades, high school grades would contribute little to explaining job performance, as was found. On the other hand, the subject matter content of courses in graduate school (say, a master's program in business) is likely to be considerably different from that of undergraduate courses, especially in the case of persons with an undergraduate degree in engineering. Thus we might expect to observe an additive effect on job performance, as was found (Table 1). It may be seen (Table 3) that the effect is stronger for engineering undergraduates than for business majors.

---

[18] This would appear to be true of the traits which Gintis contends are important, namely, subordinacy, discipline, cognitive versus affective modes of response, and motivation according to external reward.

The emphasis on affective traits has been supported elsewhere by observing that the coefficient on level of education in most regression equations is not reduced substantially when an "ability" measure is added to the regression (see Gintis). The implication is that this coefficient would be reduced much more if measures of relevant affective traits were introduced. This argument ignores not only the effect of occupational entry requirements but also individual differences in the mastery of academic subject matter. Although the findings of this study do not minimize the importance of nonacademic traits, they suggest that affective traits do not dominate academic aptitude and knowledge in their effect on the productivity of these college graduates.

C. *College as a Filter*

We can now consider the possibility that the higher educational system selects persons with productive (preexisting) traits, but does not enhance productive ability. The argument above suggests that the relation between academic achievement and job performance is not simply due to academic aptitude or to noncognitive traits associated with doing well in school.[29] Even among persons who we know had relatively high academic aptitude when they went to college, mastery of academic subject matter was related to job performance. And we have argued that the relationship seems not to be explained by preexisting affective traits which may be associated with doing well in school, given academic aptitude or intelligence. The evidence suggests that college education is not only a signal of productive ability, but in fact enhances this ability.

D. *Preassignment to Jobs with a Future*

Finally, it might be argued that persons from better schools and with better grades were initially assigned to positions with greater possibilities for advancement relative to their productive abilities than were persons with poorer grades and from lower quality colleges. The limited evidence available, however, does not lend much support to this possibility. It seems reasonable to assume that positive responses to the questions regarding initial job experience and supervisor tend to reflect a favorable initial position. According to the above hypothesis, a positive relation between college selectivity (or grades) and the initial supervisor index (or initial experience index) would be expected. But it may be seen in the tabulation below that the relevant correlations are quite small; one in fact is negative.

	SEL	*GPA*
INT	.065	−.012
SUP	.075	.032

IV. Conclusions and Comparison with Other Studies

The hypothesis of no relation between academic achievement and job performance must be rejected for this sample of college graduates. The differences in estimated rates of salary increase by college selectivity and college grade point average are substantial. These findings are consistent with those of Weisbrod and Peter Karpoff, who related earnings to college class rank and college quality using a sample of employees of the American Telephone and Telegraph Company.[30] Although they were unable to control for other individual characteristics, their calculated salary index increases with both class rank and college quality.[31] If persons with the same level of education can be distinguished by the "quality" of their degrees, then surely there would be an even greater difference in job performance between persons with different levels of education; for example, high school versus college. The measures used in selection and certification are not only related to productive traits, but the evidence suggests that college education enhanced the productive capabilities of these persons. It has also been found that other individual characteristics seem just as important as academic attributes in the determination of job productivity.

---

[29] We can rule out consideration of productive traits not associated with academic achievement on the basis of the observation that persons who do better in school also do better on the job. There seems to be no reason to believe that colleges select persons with productive traits that are not related to academic achievement. For example, traits that may be associated with socioeconomic background appear to be unimportant when academic achievement is controlled for.

[30] The data used represented within-group means

Although not affording a direct comparison, the conclusions reached by the industrial psychologist Ghiselli, who used an entirely different approach, tend to parallel the findings here, especially with respect to the nonacademic variables. From studies of managers, he concluded that the following traits were of particular importance to managerial success: intelligence, supervisory ability, initiative, self-assurance, and perceived occupational level. In drawing comparisons with this study, it might be assumed that intelligence is loosely captured (along with knowledge) by college selectivity and grades and that supervisory ability is represented by the leadership index. Initiative, which he described as a "certain independence and inventiveness," has been assumed to be partially reflected in the initial experience (or initiative) index. Self-assurance, which he assumed to be possessed by an individual who perceives himself as being effective in dealing with problems or, who is "self confident," is likely to be inversely related to the job security index.[32] Finally, perceived occupational level, which may be a "level of aspiration," would presumably be affected by socioeconomic background, although this index was not significant in this study after controlling for other variables.

Without discussing the merits of research methodologies, some indication of the reasons for the apparent discrepancy between the conclusions of persons studying blue collar workers (for example, Berg) and the results suggested by this study is provided by Wolfle and Joseph Smith. They found that high school class rank—which might be taken as a measure of academic ability or achievement—had little to do with earnings twenty years later for persons who entered technical schools, had some college, or had no post-high school education, but was strongly related to earnings for those with one or more college degrees. Results pertaining to jobs not normally held by college graduates are likely to be considerably different from those reported here.

In conclusion, measures of academic achievement and ability used in the selection and certification process in higher education are not only related to the productivity of college graduates in this sample, but the results suggest that college education contributed to their productive ability. These findings lend support to the practice of selecting students on the basis of academic measures. But nonacademic attributes, largely independent of academic characteristics, have also been shown to affect productivity. The two groups seem to be of approximately equal

---

rather than individual observations, where the groups were defined by class rank, college quality, and years employed. College quality was defined by the "subjective assessment" of company personnel officers.

[31] A similar pattern was obtained in this study using the "salary index" but the results are not shown here.

[32] Ghiselli in fact says that: "Self-realization and autonomy universally are more important to managers than prestige, social satisfaction, and even security" (p. 237).

importance. In light of the use of the college degree as an occupational screening device, this suggests a second look at the practice of selecting persons for higher education solely or largely on the basis of academic aptitude or achievement. If persons were selected for higher education on the basis of their potential productivity in a chosen occupation, rather than their potential as future students, consideration of nonacademic as well as academic attributes would be necessary.

REFERENCES

K. J. Arrow, "Higher Education as a Filter," Inst. for Math. Stud. in the Soc. Sciences, tech. rep. no. 71, Stanford Univ., Sept. 1972.

O. Ashenfelter and J. D. Mooney, "Graduate Education, Ability, and Earnings," *Rev. Econ. Statist.*, Feb. 1968, *50*, 78–86.

A. W. Astin, (1965a) "College Preferences of Very Able Students," *Coll. and Univ.*, Spring 1965, *40*, 282–97.

———, (1965b) *Who Goes Where to College*, Chicago 1965.

———, "Undergraduate Institutions and the Production of Scientists," *Science*, July 26, 1963, *141*, 334–45.

———, *Predicting Academic Performance in College*, New York 1971.

L. J. Baird and J. L. Holland, "The Flow of High School Students to Schools, Colleges, and Jobs," American College Testing Program res. rep. no. 26, June 1968.

——— and J. M. Richards, Jr., "The Effects of Selecting College Students by Various Kinds of High School Achievement," American College Testing Program res. rept. no. 23, Feb. 1968.

C. J. Bajema, "A Note on the Interrelations Among Intellectual Ability, Educational Attainment, and Occupational Achievement: A Follow-Up Study of a Male Kalamazoo Public School Population," *Soc. Educ.*, Summer 1968, *41*, 317–19.

G. S. Becker, *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*, New York 1964.

I. Berg, *Education and Jobs: The Great Training Robbery*, New York 1970.

B. S. Bloom, *Taxonomy of Educational Objectives; the Classification of Educational Goals: Handbook I: Cognitive Domain*, New York 1956.

———, D. R. Krathwohl, and B. B. Masia, *Taxonomy of Educational Objectives; The Classification of Educational Goals: Handbook II: Affective Domain*, New York 1964.

O. D. Duncan, "Ability and Achievement," *Eugen. Quart.*, Mar. 1968, *15*, 1–11.

J. C. Flanagan et al., "The American High School Student," tech. rep. to U.S. Office of Education, coop. res. proj. no. 635, Univ. Pittsburgh, Washington 1964.

E. F. Friedenberg, "Effects of the Technological Mystique on Schooling," in S. G. Tickton, ed., *To Improve Learning: An Evaluation of Educational Technology*, New York 1971.

J. W. Getzels and P. W. Jackson, *Creativity and Intelligence*, New York 1961.

H. Gintis, "Education, Technology, and the Characteristics of Worker Productivity," *Amer. Econ. Rev. Proc.*, May 1971, *41*, 266–79.

E. E. Ghiselli, "Managerial Talent," in D. Wolfle, ed., *The Discovery of a Talent*, Cambridge, Mass. 1969.

D. A. Goslin, *The Search for Ability: Standardized Testing in Social Perspective*, New York 1963.

———, "Standardized Ability Tests and Testing: Major Issues and the Validity of Current Criticisms of Tests are Discussed," *Science*, Feb. 23, 1968, *159*, 851–55.

Z. Griliches and W. M. Mason, "Education, Income, and Ability," *J. Polit. Econ.*, May/June 1972, suppl., *80*, S74–S103.

G. Hanoch, "An Economic Analysis of Earnings and Schooling," *J. Hum. Resources*, Summer 1967, *2*, 310–29.

L. W. Hansen, B. A. Weisbrod, and W. J. Scanlon, "Schooling and Earnings of Low Achievers," *Amer. Econ. Rev.*, June 1970, *60*, 409–18.

J. L. Holland and R. C. Nichols, "Prediction of Academic and Extracurricular Achievement in College," *J. Educ. Psychology*, Feb. 1964, *55*, 55–65.

——— and J. M. Richards, Jr., "Academic and Non-Academic Accomplishment: Correlated or Uncorrelated," American College Testing Program res. rep. no. 2, Apr. 1965.

——— and ———, "Academic and Non-Academic Accomplishment in a Representative Sample Taken from a Population of 612,000," American College Testing Program res. rep. no. 12, May 1966.

D. P. Hoyt, "The Relationship Between College Grades and Adult Achievement: A Review of the Literature," American College Testing Program res. rep. no. 7, Sept. 1965.

D. W. MacKinnon, "The Nature and Nurture of Creative Talent," *Amer. Psychologist*, July 1962, *17*, 484–95.

J. Mincer, "On-the-Job Training," *J. Polit. Econ.*, Oct. 1962, suppl., *70*, 50–79.

S. Payne, "Good Scholars Not Always Best," *Bus. Week*, Feb. 24, 1962, 77–78.

J. P. Robinson, R. Athanasiou, and K. B. Head, "Measures of Occupational Attitudes and Occupational Characteristics," Surv. Res. Center, Univ. Michigan 1969.

M. Spence, "Market Signalling: The Informational Structure of Job Markets and Related Phenomena," Public Policy Program disc. pap. no. 4D, Harvard Univ., Feb. 1972.

P. J. Taubman and T. J. Wales, "Higher Education, Mental Ability, and Screening," *J. Polit. Econ.*, Jan./Feb. 1973, *81*, 28–55.

C. W. Taylor and R. L. Ellison, "Biographical Predictors of Scientific Performance," *Science*, Mar. 3, 1967, *155*, 1075–80.

L. C. Thurow, "The Occupational Distribution of the Returns to Education and Experience for Whites and Negroes," *1967 Proc. Amer. Statist. Assn., Soc. Statist. Sec.*, Washington 1968, 233–44.

———, *Investment in Human Capital*, Belmont 1970.

B. A. Weisbrod and P. Karpoff, "Monetary Returns to College Education, Student Ability, and College Quality," *Rev. Econ. Statist.*, Nov. 1968, *50*, 491–97.

Y. Weiss, "Investment in Graduate Education," *Amer. Econ. Rev.*, Dec. 1971, *61*, 833–52.

M. R. Winterbottom, "The Relationship of Childhood Training in Independence to Achievement Motivation," in D. C. McClelland et al., eds., *The Achievement Motive*, New York 1953, 297–304.

D. A. Wise, "Academic Achievement and Job Performance: Earnings and Promotions," Ford Foundation Program Res. Univ. Admin., pap. P-35, Univ. California, Berkeley, 1972.

———, "Personal Attributes, Job Performances, and Probability of Promotion," *Econometrica*, forthcoming 1975.

D. Wolfle, "The Freshman Class," *Science*, Sept. 24, 1965; *149*, p. 1453.

——— **and J. G. Smith**, "The Occupational Value of Education for Superior High School Graduates," *J. Higher Educ.*, Apr. 1956, *27*, 201–12.

College Entrance Examination Board (*CEEB*), *Report of the Commission on Tests: I. Righting the Balance*, Princeton 1970.

Student in Higher Education, Report of the Committee on the Student in Higher Education, The Hazen Foundation, New Haven 1968.

Transcendental Logarithmic Utility Functions

By LAURITS R. CHRISTENSEN, DALE W. JORGENSON, AND LAWRENCE J. LAU*

The traditional starting point for econometric studies of consumer demand is a system of demand functions giving the quantity consumed of each commodity as a function of total expenditure and the prices of all commodities. Tests of the theory of demand are formulated by requiring that the demand functions be consistent with utility maximization. Additive and homothetic utility functions have played an important role in formulating tests of the theory of demand. If the utility function is homothetic, expenditure proportions are independent of total expenditure. If the utility function is additive and homothetic, elasticities of substitution among all pairs of commodities are constant and equal.[1]

An example of the traditional approach to demand analysis is the system of double logarithmic demand functions employed in the pioneering studies of consumer demand by Henry Schultz, Richard Stone, and Herman Wold. If the theory of demand is valid and demand functions are double logarithmic, the utility function is linear logarithmic.[2] Similarly, the Rotterdam system of demand functions employed by A. P. Barten and Henri Theil is consistent with utility maximization only if the utility function is linear logarithmic.[3] A linear logarithmic utility function is both additive and homothetic; all expenditure proportions are constant, and elasticities of substitution among all pairs of commodities are constant and equal to unity.

Hendrik Houthakker and Stone have developed alternative approaches to demand analysis that retain the assumption of additivity while dropping the assumption of homotheticity.[4] Stone has employed a linear expenditure system, based on a utility function that is linear in the logarithms of quantity consumed less a constant for each commodity. The constants are interpreted as initial commitments; incremental expenditure proportions, derived from quantities consumed in excess of the initial commitments, are constant for all variations in total expenditure and prices. If all initial commitments are zero, the utility function is linear logarithmic in form. Nonzero commitments permit expenditure proportions to vary with total expenditure.

Houthakker has employed a direct addilog system, based on a utility function that is additive in functions that are homogeneous in the quantity consumed for each commodity. The degree of homogeneity may differ from commodity to commodity, permitting expenditure proportions to vary with total expenditure. If the degree of homogeneity is the same for all commodities, the addilog utility function is additive and homothetic. Robert Basmann, Leif Johansen, and Kazuo Sato have combined the approaches of Houthakker and Stone, defining each of the homogeneous functions in the direct addilog

---

* University of Wisconsin-Madison, Harvard University, and Stanford University, respectively.

[1] The class of additive and homothetic utility functions was first characterized by Abram Bergson.

[2] See Schultz, Stone (1954a), Wold, and Robert Basmann, R. C. Battalio, and J. H. Kagel.

[3] See Barten (1964, 1967, 1969), Daniel McFadden, and Theil (1965, 1967, 1971).

[4] See Houthakker (1960) and Stone (1954b). The linear expenditure system was originally proposed by Lawrence Klein and Herman Rubin.

utility function on the quantity consumed less a constant for each commodity. The resulting utility function is additive but not homothetic.[5]

Our first objective is to develop tests of the theory of demand that do not employ additivity or homotheticity as part of the maintained hypothesis. For this purpose we introduce new representations of the utility function in Section I. Our approach is to represent the utility function by functions that are quadratic in the logarithms of the quantities consumed. The resulting utility functions provide a local second-order approximation to any utility function. These utility functions allow expenditure shares to vary with the level of total expenditure and permit a greater variety of substitution patterns among commodities than functions based on constant and equal elasticities of substitution among all pairs of commodities.

Our second objective is to exploit the duality between prices and quantities in the theory of demand. A complete model of consumer demand implies the existence of an indirect utility function, defined on total expenditure and the prices of all commodities.[6] The indirect utility function is homogeneous of degree zero and can be expressed as a function of the ratios of prices of all commodities to total expenditure. The indirect utility function is useful in characterizing systems of direct demand functions, giving quantities consumed as functions of the ratios of prices to total expenditure. The direct utility function is useful in characterizing systems of indirect demand functions, giving the ratios of prices to total expenditure as functions of the quantities consumed. The system consisting of direct utility function and indirect demand functions is dual to the system consisting of indirect utility function and direct demand functions.[7]

We represent the indirect utility function by functions that are quadratic in the logarithms of ratios of prices to total expenditure, paralleling our treatment of the direct utility function. The resulting indirect utility functions provide a local second-order approximation to any indirect utility function. These indirect utility functions are not required to be additive or homothetic. The duality between direct and indirect utility functions has been used extensively in Houthakker's pathbreaking studies of consumer demand. Paralleling the direct addilog demand system, Houthakker has employed an indirect addilog system, based on an indirect utility function that is additive in ratios of prices to total expenditure.[8]

We refer to our representation of the direct utility function as the direct transcendental logarithmic utility function, or more simply, the *direct translog utility function*. The utility function is a transcendental function of the logarithms of quantities consumed. Similarly we refer to our representation of the indirect utility function as the indirect transcendental logarithmic utility function, or, more simply, the *indirect translog utility function*. Earlier, we introduced transcendental logarithmic functions into the study of production.[9] The duality between direct and indirect translog utility functions is analogous to the duality between translog production and price frontiers employed in our study of production.

---

[5] A recent survey of econometric studies of consumer demand is given by Alan Brown and Angus Deaton.

[6] The indirect utility function was introduced by Giovanni Antonelli, and independently by Harold Hotelling.

[7] Indirect demand functions were introduced by Antonelli. The duality between direct and indirect utility functions is discussed by Lau (1969a); the duality between systems of direct and indirect demand functions is discussed by Wold, John Chipman, and Leonid Hurwicz

[8] See Houthakker (1960). This demand system was originally proposed by Conrad Leser and has also been employed by W. H. Somermeyer, J. G. M. Hilhorst, and J. W. W. A. Wit.

[9] See the authors (1971, 1973).

For an additive direct utility function ratios of indirect demand functions, giving the ratios of prices, depend only on the quantities of the two commodities involved. The direct addilog and linear expenditure systems, together with the system employed by Basmann, Johansen, and Sato, have this property. Similarly, for an additive indirect utility function, ratios of direct demand functions giving the ratios of quantities depend only on the prices of the two commodities involved. The indirect addilog system has this property. For an additive and homothetic direct utility function the ratios of indirect demand functions depend only on the ratios of quantities. Furthermore, the indirect utility function is also additive and homothetic, so that ratios of direct demand functions depend only on ratios of prices. The use of direct and indirect translog utility functions permits us to test these restrictions on direct and indirect demand functions. We do not impose the restrictions as part of the maintained hypothesis.

We present statistical tests of the theory of demand in Section II. These tests can be divided into two groups. First, we test restrictions on the parameters of the direct translog utility function implied by the theory of demand. We test these restrictions without imposing the assumptions of additivity and homotheticity. We test precisely analogous restrictions on the parameters of the indirect translog utility function. Second, we test restrictions on the direct translog utility function corresponding to restrictions on the form of the utility function. In particular, we test restrictions corresponding to additivity and homotheticity of the direct translog utility function. Again, we test precisely analogous restrictions on the indirect translog utility function.

We present empirical tests of the theory of demand based on time-series data for the United States for 1929–72 in Section III. The data include prices and quantities consumed of the services of consumers' durables, nondurable goods, and other services. For these data we present direct tests of the theory of demand based on the direct translog utility function, and indirect tests of the theory based on the indirect translog utility function. For both direct and indirect tests we first test the extensive set of restrictions implied by the theory of demand. Proceeding conditionally on the validity of the theory, we test restrictions on the form of the direct and indirect utility functions implied by the assumptions of additivity and homotheticity.

I. Transcendental Logarithmic Utility Functions

A. *Direct Translog Utility Function*

The ***direct utility function U*** can be represented in the form:

$$\ln U = \ln U(X_1, X_2, \ldots, X_m) \tag{1}$$

where X_i is the quantity consumed of the ith commodity. The consumer maximizes utility subject to the budget constraint:

$$\sum p_i X_i = M \tag{2}$$

where p_i is the price of the ith commodity and M is the value of total expenditure.

The first-order conditions for a maximum of utility can be written:

$$\frac{\partial \ln U}{\partial \ln X_j} = \mu \frac{p_j X_j}{U} \quad (j = 1, 2, \ldots, m) \tag{3}$$

where μ is the marginal utility of income. From the budget constraint we obtain:

$$\frac{\mu}{U} = \frac{1}{M} \sum \frac{\partial \ln U}{\partial \ln X_i} \tag{4}$$

so that:

$$\frac{\partial \ln U}{\partial \ln X_j} = \frac{p_j X_j}{M} \sum \frac{\partial \ln U}{\partial \ln X_i} \quad (j = 1, 2, \ldots, m) \tag{5}$$

To preserve symmetry with our treatment of the indirect utility function given below, we approximate the negative of the logarithm of the direct utility function by a function quadratic in the logarithms of the quantities consumed:

$$(6)\quad -\ln U = \alpha_0 + \sum \alpha_i \ln X_i + \tfrac{1}{2}\sum\sum \beta_{ij} \ln X_i \ln X_j$$

Using this form for the utility function we obtain:

$$(7)\quad \alpha_j + \sum \beta_{ji} \ln X_i = \frac{p_j X_j}{M} \sum (\alpha_k + \sum \beta_{ki} \ln X_i) \qquad (j = 1, 2, \ldots, m)$$

To simplify notation we write:

$$(8)\quad \alpha_M = \sum \alpha_k$$

$$(9)\quad \beta_{Mi} = \sum \beta_{ki} \qquad (i = 1, 2, \ldots, m)$$

so that:

$$(10)\quad \frac{p_j X_j}{M} = \frac{\alpha_j + \sum \beta_{ji} \ln X_i}{\alpha_M + \sum \beta_{Mi} \ln X_i} \qquad (j = 1, 2, \ldots, m)$$

The budget constraint implies that:

$$(11)\quad \sum \frac{p_i X_i}{M} = 1$$

so that, given the parameters of any $m-1$ equations for the budget shares $p_j X_j / M$ $(j=1, 2, \ldots, m)$, the parameters of the mth equation, α_m and β_{mj} $(j=1, 2, \ldots, m)$, can be determined from the definitions of α_M and β_{Mj} $(j=1, 2, \ldots, m)$.

Since the equations for the budget shares are homogeneous of degree zero in the parameters, a normalization of the parameters is required for estimation. A convenient normalization for the parameters of the direct translog utility function is:

$$(12)\quad \alpha_M = \sum \alpha_i = -1$$

B. *Indirect Translog Utility Function*

The ***indirect utility function*** V can be represented in the form:

$$(13)\quad \ln V = \ln V\left(\frac{p_1}{M}, \frac{p_2}{M}, \ldots, \frac{p_m}{M}\right)$$

We determine the budget share for the jth commodity from the identity:[10]

$$(14)\quad \frac{p_j X_j}{M} = -\frac{\partial \ln V}{\partial \ln p_j} \Big/ \frac{\partial \ln V}{\partial \ln M} \qquad (j = 1, 2, \ldots, m)$$

Preserving symmetry with the direct utility function, we approximate the logarithm of the indirect utility function by a function quadratic in the logarithms of the ratios of prices to the value of total expenditure:

$$(15)\quad \ln V = \alpha_0 + \sum \alpha_i \ln \frac{p_i}{M} + \tfrac{1}{2}\sum\sum \beta_{ij} \ln \frac{p_i}{M} \ln \frac{p_j}{M}$$

Using this form for the utility function we obtain:

$$(16)\quad \frac{\partial \ln V}{\partial \ln p_j} = \alpha_j + \sum \beta_{ji} \ln \frac{p_i}{M} \qquad (j = 1, 2, \ldots, m)$$

$$(17)\quad -\frac{\partial \ln V}{\partial \ln M} = \sum\left(\alpha_k + \sum \beta_{ki} \ln \frac{p_i}{M}\right)$$

As before, we simplify notation by writing:

$$(18)\quad \alpha_M = \sum \alpha_k$$

$$(19)\quad \beta_{Mi} = \sum \beta_{ki} \qquad (i = 1, 2, \ldots, m)$$

so that:

$$(20)\quad \frac{p_j X_j}{M} = \frac{\alpha_j + \sum \beta_{ji} \ln \frac{p_i}{M}}{\alpha_M + \sum \beta_{Mi} \ln \frac{p_i}{M}} \qquad (j = 1, 2, \ldots, m)$$

---

[10] This is the logarithmic form of René Roy's Identity.

The budget constraint implies that given the parameters of any $m-1$ equations for the budget shares, the parameters of the mth equation, α_m and β_{mj} $(j=1, 2, \dots, m)$, can be determined from the definitions of α_M and β_{Mj} $(j=1, 2, \dots, m)$. As before, we can normalize the parameters so that:

$$\alpha_M = \sum \alpha_i = -1 \tag{21}$$

II. Testing the Theory of Demand

A. *Stochastic Specification*

The first step in implementing an econometric model of demand based on the direct translog utility function is to add a stochastic specification to the theoretical model based on equations for the budget shares p_jX_j/M $(j=1, 2, \dots, m)$. Given the disturbances in any $m-1$ equations, the disturbance of the remaining equation can be determined from the budget constraint. Only $m-1$ equations are required for a complete econometric model of demand.

B. *Equality and Symmetry*

We estimate $m-1$ equations for the budget shares, subject to normalization of the parameter α_M appearing in each equation at minus unity. If the equations are generated by utility maximization, the parameters β_{Mj} $(j=1, 2, \dots, m)$ appearing in each equation must be the same. This results in a set of restrictions relating the m parameters appearing in each of the $m-1$ equations, a total of $m(m-2)$ restrictions. We refer to these as *equality restrictions.*

The logarithm of the direct translog utility function is twice differentiable in the logarithms of the quantities consumed, so that the Hessian of this function is symmetric. This gives rise to a set of restrictions relating the parameters of the cross-partial derivatives:

$$\beta_{ij} = \beta_{ji} \quad (i \neq j;\ i, j = 1, 2, \dots, m) \tag{22}$$

There are $(1/2)(m-2)(m-1)$ restrictions of this type among the parameters of the $m-1$ equations we estimate directly and $m-1$ such restrictions among the parameters of the equations we estimate indirectly from the budget constraint. We refer to these as *symmetry restrictions.* The total number of symmetry restrictions is $(1/2)m(m-1)$.

If equations for the budget shares are generated by maximization of a direct translog utility function, the parameters satisfy equality and symmetry restrictions. There are $(1/2)m(3m-5)$ such restrictions.

C. *Additivity*

If the direct utility function U is *additive,* we can write

$$\ln U = F\left(\sum \ln U^i(X_i)\right) \tag{23}$$

where each of the functions U^i depends on only one of the commodities demanded X_i, and F is a real-valued function of one variable.

The parameters of the translog approximation to an additive direct utility function can be written:

$$-\frac{\partial \ln U}{\partial \ln X_i} = -F'\frac{\partial \ln U^i}{\partial \ln X_i} = \alpha_i \quad (i = 1, 2, \dots, m) \tag{24}$$

$$-\frac{\partial^2 \ln U}{\partial \ln X_i \partial \ln X_j} = -F''\frac{\partial \ln U^i}{\partial \ln X_i}\frac{\partial \ln U^j}{\partial \ln X_j} = \beta_{ij} \quad (k \neq j;\ i, j = 1, 2, \dots, m) \tag{25}$$

where the logarithmic derivatives,

$$F' = \frac{\partial \ln U}{\partial \ln U^i} \quad (i = 1, 2, \dots, m) \tag{26}$$

$$F'' = \frac{\partial^2 \ln U}{\partial \ln U^i \partial \ln U^j} \quad (i, j = 1, 2, \dots, m) \tag{27}$$

are independent of i and j.

Under additivity the parameters of the translog utility function satisfy the restrictions:

$$\beta_{ij} = \theta\alpha_i\alpha_j \qquad (i \neq j;\ i, j = 1, 2, \ldots, m) \tag{28}$$

where

$$\theta = -\frac{F''}{(F')^2} \tag{29}$$

We refer to these as *additivity restrictions.* The total number of such restrictions is $(1/2)(m-2)(m-1)$.

The translog approximation to an additive utility function is not necessarily additive. The direct translog utility function is additive if and only if $ln\ U$ can be written as the sum of m functions, each depending on only one of the quantities demanded. We refer to such a function as *explicitly additive.* Explicit additivity of the translog utility function implies the additivity restrictions given above and the additional restriction:

$$\theta = 0 \tag{30}$$

We refer to this restriction as the *explicit additivity restriction.* We note that the translog approximation to an explicitly additive function is explicitly additive.

D. *Homotheticity*

If the direct utility function is *homothetic,* we can write:

$$ln\ U = F(ln\ H(X_1, X_2, \ldots, X_m)) \tag{31}$$

where the function H is homogeneous of degree one.

The parameters of the translog approximation to a homothetic direct utility function can be written:

$$-\frac{\partial\, ln\ U}{\partial\, ln\ X_i} = \frac{\partial F}{\partial\, ln\ H}\, \frac{\partial\, ln\ H}{\partial\, ln\ X_i} = \alpha_i \qquad (i = 1, 2, \ldots, m) \tag{32}$$

$$-\frac{\partial^2\, ln\ U}{\partial\, ln\ X_i \partial\, ln\ X_j} = \left[\frac{\partial F}{\partial\, ln\ H}\, \frac{\partial^2\, ln\ H}{\partial\, ln\ X_i \partial\, ln\ X_j} + \frac{\partial^2 F}{\partial\, ln\ H^2}\, \frac{\partial\, ln\ H}{\partial\, ln\ X_i}\, \frac{\partial\, ln\ H}{\partial\, ln\ X_j}\right] = \beta_{ij} \qquad (i, j = 1, 2, \ldots, m) \tag{33}$$

Homogeneity of degree one of the function H implies that:

$$\sum \frac{\partial\, ln\ H}{\partial\, ln\ X_i} = 1 \tag{34}$$

$$\sum_{j=1}^{m} \frac{\partial^2\, ln\ H}{\partial\, ln\ X_i \partial\, ln\ X_j} = 0 \qquad (i = 1, 2, \ldots, m) \tag{35}$$

Under homotheticity the parameters of the translog utility function satisfy the restrictions:

$$\beta_{Mi} = \sigma\alpha_i \qquad (i = 1, 2, \ldots, m) \tag{36}$$

where

$$\sigma = \frac{\partial^2 F}{\partial\, ln\ H^2} \tag{37}$$

and we have used the normalization

$$\sum \alpha_i = -1 \tag{38}$$

We refer to these as *homotheticity restrictions.* There are $m-1$ such restrictions.

The translog approximation to a homothetic utility function is not necessarily homothetic. A necessary and sufficient condition for homotheticity of the translog utility function is that it is homogeneous, so that:

$$\sigma = 0 \tag{39}$$

We refer to this restriction as the *homogeneity restriction.* We note that the translog approximation to a homogeneous function is homogeneous.

E. *Additivity and Homotheticity*

The class of additive and homothetic direct utility functions coincides with the class of utility functions with constant elasticities of substitution among all commodities. If the direct utility function U is additive and homothetic, we can write:

$$\ln U = F\left(\sum \delta_i X_i^\rho\right) \tag{40}$$

or

$$\ln U = F\left(\sum \delta_i \ln X_i\right) \tag{41}$$

The second form is a limiting case of the first corresponding to unitary elasticities of substitution among all commodities.

The parameters of the translog approximation to an additive and homothetic utility function satisfy the additivity and homotheticity restrictions given above, so that:

$$\beta_{ij} = \theta\alpha_i\alpha_j \quad (i \neq j;\, i,j = 1, 2, \ldots, m) \tag{42}$$

$$\beta_{ii} = (\sigma + \theta)\alpha_i + \theta\alpha_i^2 \quad (i = 1, 2, \ldots, m) \tag{43}$$

where we have used the normalization

$$\sum \alpha_i = -1 \tag{44}$$

The translog approximation to an additive and homothetic utility function is not necessarily additive and homothetic. We can, however, identify the parameters of a translog approximation with the parameters of the additive and homothetic utility function given above, as follows:

$$\sigma + \theta = \rho \tag{45}$$

$$\alpha_i = \delta_i \quad (i = 1, 2, \ldots, m) \tag{46}$$

As before, the parameter θ corresponds to:

$$\theta = -\frac{F''}{(F')^2} \tag{47}$$

The translog approximation to a utility function characterized by unitary elasticities of substitution among all commodities satisfies the restriction:

$$\sigma + \theta = \rho = 0 \tag{48}$$

Under this restriction the expenditures shares are constant:

$$\frac{p_j X_j}{M} = \frac{\alpha_j + \theta\alpha_j \sum \alpha_i \ln X_i}{-1 - \theta \sum \alpha_i \ln X_i} = -\alpha_j \tag{49}$$

The value of the parameter θ is arbitrary so that we can let σ and θ be equal to zero. The translog approximation to a utility function with unitary elasticities of substitution has the same empirical implications as a linear logarithmic utility function, which is explicitly additive and homogeneous.

F. *Duality*

In implementing an econometric model of demand based on the indirect translog utility function the first step as before is to add a stochastic specification to the theoretical model based on equations for the budget shares $p_j X_j/M$ $(j=1, 2, \ldots, m)$. Only $m-1$ equations are required for a complete model. If equations for the budget shares are generated from the indirect translog utility function, the parameters satisfy equality and symmetry restrictions that are strictly analogous to the corresponding restrictions for the direct translog utility function.

Additivity and homotheticity restrictions for the indirect translog utility function are analogous to the corresponding restrictions for the direct translog function. The direct utility function is homothetic if and only if the indirect utility function is homothetic.[11] In general, an additive direct utility function does not correspond to an additive indirect utility function.[12] However, if the direct utility function is additive and homothetic, the indirect util-

[11] See, for instance, Paul Samuelson (1965) and Lau (1969b).

[12] See Lau (1969b).

ity function is additive and homothetic. Direct and indirect utility functions corresponding to the same preferences are additive only if the direct utility function is homothetic[13] or the direct utility function is linear logarithmic in all but one of the commodities.[14] In this relationship the roles of direct and indirect utility functions can be interchanged.

A direct utility function is *self-dual* if the corresponding indirect utility function has the same form.[15] The only direct translog utility function which is self-dual is the linear logarithmic utility function. Linear logarithmic utility functions are the only additive and homothetic direct or indirect translog utility functions. Translog direct and indirect utility functions represent the same preferences if and only if they are self-dual; that is, if and only if they are linear logarithmic. Unless this stringent condition is met, the direct and indirect translog approximations to a given pair of direct and indirect utility functions represent different preferences.

III. Empirical Results

A. *Summary of Tests*

We have developed statistical tests of the theory of demand that do not employ the assumptions of additivity and homotheticity for individual commodities or for commodity groups. At this point we outline restrictions on equations for the budget shares corresponding to direct and indirect translog utility functions. First, we present restrictions derived from the theory of demand; that is, from the basic hypothesis of utility maximization. Second, we present restrictions derived from hypotheses about functional form such as additivity and homotheticity.

Our empirical tests are based on data for three commodity groups—services of consumers' durables, nondurable goods, and other services—so that we specialize our presentation of statistical tests of the theory of demand to the three-commodity case, $m=3$. A complete econometric model for either direct or indirect translog utility functions is provided by any pair of equations for the budget shares. We consider the system of two equations,

$$\frac{p_1X_1}{M} = \frac{\alpha_1+\beta_{11}\, ln\, X_1+\beta_{12}\, ln\, X_2+\beta_{13}\, ln\, X_3}{-1+\beta_{M1}\, ln\, X_1+\beta_{M2}\, ln\, X_2+\beta_{M3}\, ln\, X_3} \tag{50}$$

$$\frac{p_2X_2}{M} = \frac{\alpha_2+\beta_{21}\, ln\, X_1+\beta_{22}\, ln\, X_2+\beta_{23}\, ln\, X_3}{-1+\beta_{M1}\, ln\, X_1+\beta_{M2}\, ln\, X_2+\beta_{M3}\, ln\, X_3} \tag{51}$$

corresponding to the direct translog utility function and the system of equations,

$$\frac{p_1X_1}{M} = \frac{\alpha_1+\beta_{11}\, ln\, \frac{p_1}{M}+\beta_{12}\, ln\, \frac{p_2}{M}+\beta_{13}\, ln\, \frac{p_3}{M}}{-1+\beta_{M1}\, ln\, \frac{p_1}{M}+\beta_{M2}\, ln\, \frac{p_2}{M}+\beta_{M3}\, ln\, \frac{p_3}{M}} \tag{52}$$

$$\frac{p_2X_2}{M} = \frac{\alpha_2+\beta_{21}\, ln\, \frac{p_1}{M}+\beta_{22}\, ln\, \frac{p_2}{M}+\beta_{23}\, ln\, \frac{p_3}{M}}{-1+\beta_{M1}\, ln\, \frac{p_1}{M}+\beta_{M2}\, ln\, \frac{p_2}{M}+\beta_{M3}\, ln\, \frac{p_3}{M}} \tag{53}$$

corresponding to the indirect translog utility function. Restrictions for the two systems are perfectly analogous so that the following outline may be applied to either system. We emphasize the fact that the two systems represent the same preferences if and only if

---

[13] See Houthakker (1960), Samuelson (1965), and Lau (1969b).

[14] This is the special case introduced by John Hicks. See also Samuelson (1969).

[15] See Samuelson (1965) and Houthakker (1965). We may also mention the "self-dual addilog system" introduced by Houthakker (1965). This system is not generated by additive utility functions except for special cases.

$$\beta_{ij} = 0 \qquad (i, j = 1, 2, 3) \tag{54}$$

1. *Equality Restrictions.* The parameters $\{\beta_{M1}, \beta_{M2}, \beta_{M3}\}$ occur in both equations and must take the same values. There are three equality restrictions.

2. *Symmetry Restrictions.* One restriction of this type is explicit in the two equations we estimate directly, namely,

$$\beta_{12} = \beta_{21} \tag{55}$$

In addition, we estimate the parameters β_{31} and β_{32} from the equations:

$$\beta_{31} = \beta_{M1} - \beta_{11} - \beta_{21} \tag{56}$$

$$\beta_{32} = \beta_{M2} - \beta_{12} - \beta_{22} \tag{57}$$

so that the two additional symmetry restrictions are implicit in the two equations we estimate. We write these restrictions in the form:

$$\beta_{13} = \beta_{M1} - \beta_{11} - \beta_{21} \tag{58}$$

$$\beta_{23} = \beta_{M2} - \beta_{12} - \beta_{22} \tag{59}$$

There are three symmetry restrictions altogether.

We can identify tests of the six equality and symmetry restrictions with tests of the theory of demand. We next consider tests of hypotheses about functional form.

3. *Additivity Restrictions.* Given the equality and symmetry restrictions, the additivity restrictions take the form:

$$\beta_{12} = \theta\alpha_1\alpha_2 \tag{60}$$

$$\beta_{13} = \theta\alpha_1(-1 - \alpha_1 - \alpha_2) \tag{61}$$

$$\beta_{23} = \theta\alpha_2(-1 - \alpha_1 - \alpha_2) \tag{62}$$

We introduce the additional parameter θ, so that there are two independent restrictions of this type.

4. *Homotheticity Restrictions.* Given the equality and symmetry restrictions, the homotheticity restrictions take the form:

$$\beta_{M1} = \sigma\alpha_1 \tag{63}$$

$$\beta_{M2} = \sigma\alpha_2 \tag{64}$$

$$\beta_{M3} = \sigma(-1 - \alpha_1 - \alpha_2) \tag{65}$$

We introduce the additional parameter σ, so that there are two independent restrictions of this type.

5. *Explicit Additivity Restrictions.* Given the equality, symmetry, and additivity restrictions, a translog utility function is explicitly additive under the further restriction:

$$\theta = 0 \tag{66}$$

6. *Homogeneity Restrictions.* Given the equality, symmetry, and homotheticity restrictions, the homogeneity restriction takes the form:

$$\sigma = 0 \tag{67}$$

7. *Linear Logarithmic Utility Restrictions.* Given the equality, symmetry, additivity, and homotheticity restrictions, the translog utility function reduces to linear logarithmic form under the additional restrictions:

$$\sigma = \theta = 0 \tag{68}$$

Tests of hypotheses about the form of the utility function can be carried out in a number of different sequences. We propose to test additivity and homotheticity in parallel. Each of these hypotheses consists of two independent restrictions. Our test procedure is presented in diagrammatic form in Figure 1. We also present alternative test procedures. For example, we could first test additivity and then test homotheticity only if additivity is accepted. Our test procedure is indicated with double lines while possible alternative procedures are indicated with single lines.

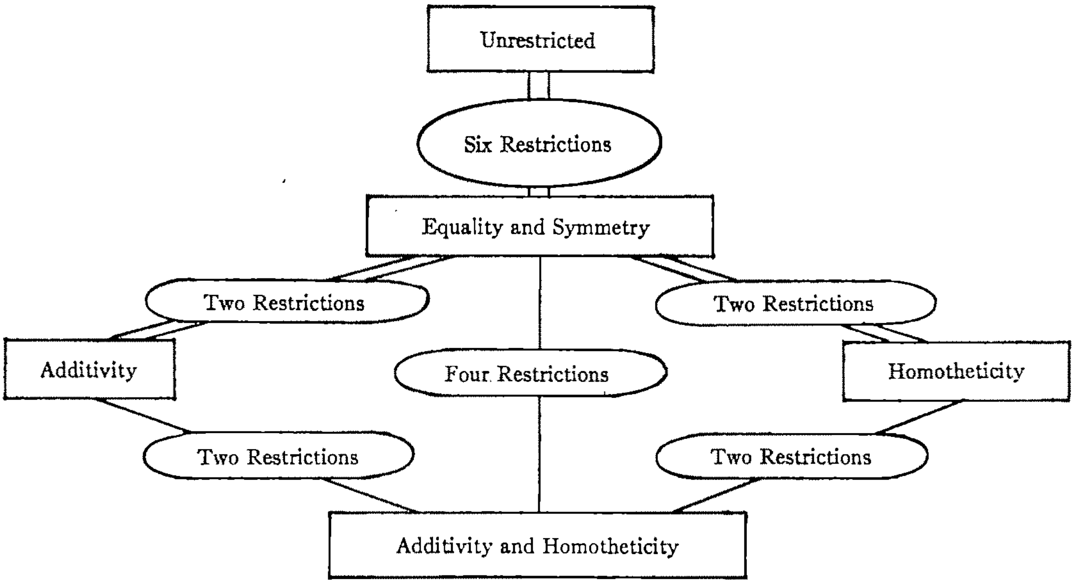


FIGURE 1. TESTS OF ADDITIVITY AND HOMOTHETICITY

The next step in our test procedure depends on the outcome of the tests of additivity and homotheticity. If we accept both additivity and homotheticity restrictions, we continue by testing explicit additivity and homogeneity in parallel. If we accept both hypotheses, we conclude that utility is linear logarithmic. A second possibility is that we accept additivity but reject homotheticity. In this case we continue by testing explicit additivity. The third possibility is that we accept homotheticity but reject additivity; we continue by testing homogeneity. If we reject both

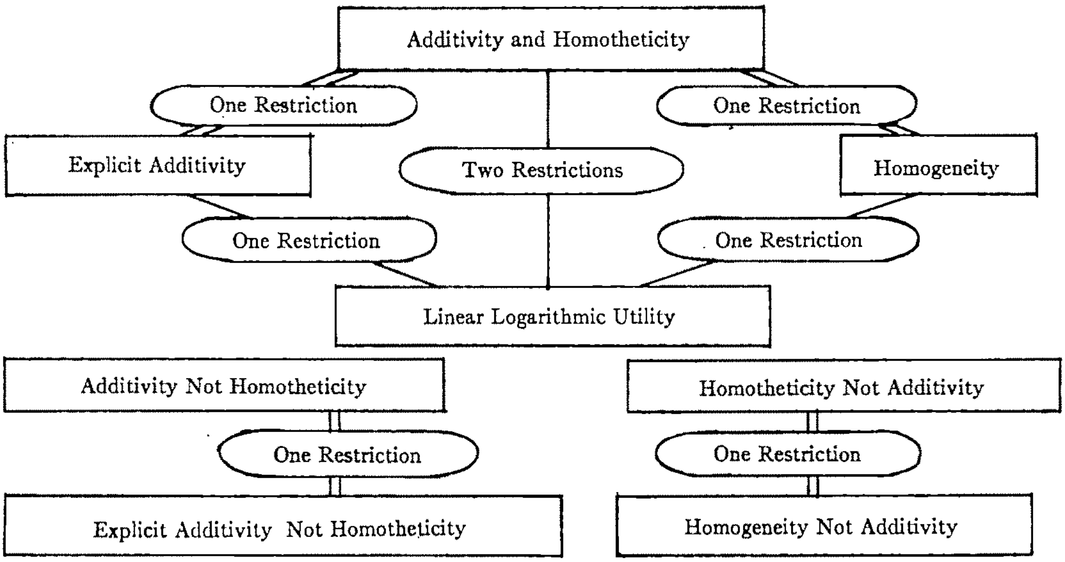

Additivity and Homotheticity
One Restriction
One Restriction
Explicit Additivity
Two Restrictions
Homogeneity
One Restriction
One Restriction
Linear Logarithmic Utility
Additivity Not Homotheticity
Homotheticity Not Additivity
One Restriction
One Restriction
Explicit Additivity Not Homotheticity
Homogeneity Not Additivity

FIGURE 2. TESTS CONDITIONAL ON ADDITIVITY OR HOMOTHETICITY

additivity and homogeneity, we terminate the sequence of tests. These tests are presented diagrammatically in Figure 2.

B. *Estimation*

Our empirical results are based on time-series data for *U. S.* personal consumption expenditures for 1929–72. The data include prices and quantities of the services of consumers' durables, nondurable goods, and other services. We have fitted the equations for budget shares generated by direct translog and indirect translog utility functions, employing the stochastic specification outlined above. Under this specification only two equations are required for a complete econometric model of demand. We have fitted equations for the services of consumers' durables (durables) and for nondurable goods (nondurables). There are forty-four observations for each behavioral equation, so that the number of degrees of freedom available for statistical tests of the theory of demand is eighty-eight for either direct or indirect specification.

For both direct and indirect specifications the maintained hypothesis consists of the unrestricted form of the two behavioral equations for the budget shares. We estimate the behavioral equations for durables and nondurables by the method of maximum likelihood and derive estimates of the parameters of the behavioral equation for services, using the budget constraint.[16] The maximum likelihood estimates of the parameters of all three behavioral equations are invariant with respect to the choice of the two equations to be estimated directly.

The unrestricted behavioral equations, normalized so that the parameter α_M is minus unity in each equation, involve fourteen unknown parameters or seven unknown parameters for each equation. Unrestricted estimates of these parameters for the direct translog utility function are presented in the first column of Table 1. The first hypothesis to be tested is that the theory of demand is valid; to test this hypothesis we impose the six equality and symmetry restrictions. Restricted estimates of the fourteen unknown parameters are presented in the second column of Table 1. Unrestricted and restricted estimates of the fourteen unknown parameters for the indirect translog utility function are presented in the first and second columns of Table 2.

Given the validity of the theory of demand, the remaining hypotheses to be tested are restrictions on the functional form. First, we test the hypothesis that the direct utility function is additive; this hypothesis requires that we impose two additional restrictions. The corresponding restricted estimates of the unknown parameters for the direct translog utility function are given in the third column of Table 1. Second, we test the hypothesis that the direct utility function is homothetic without imposing the additivity restrictions; this hypothesis requires two restrictions in addition to the equality and symmetry restrictions. The restricted estimates of the unknown parameters for the direct translog utility function are given in the fourth column of Table 1. Finally, we impose both additivity and homotheticity restrictions, obtaining the restricted estimates presented in the fifth column of Table 1. The corresponding restricted estimates for the indirect translog utility function are given in the third, fourth, and fifth columns of Table 2.

The next stage of our test procedure is contingent on the outcome of our tests of

---

[16] We employ the maximum likelihood estimator discussed, for example, by Edmond Malinvaud, pp. 338–41. In the computations we use the Gauss-Newton method described by Malinvaud, p. 343. For the direct series of tests we assume that the disturbances are independent of the quantities consumed. For the indirect series of tests we assume that the disturbances are independent of the ratios of prices to the value of total expenditure.

TABLE 1—ESTIMATES OF THE PARAMETERS OF THE DIRECT TRANSLOG UTILITY FUNCTION

Parameters	Unrestricted Estimates	Equality and Symmetry	Additivity	Homotheticity	Additivity and Homotheticity	Explicit Additivity, Not Homotheticity	Homogeneity, Not Additivity	Explicit Additivity and Homotheticity	Additivity and Homogeneity	Linear Logarithmic Utility
					DURABLES					
α_1	−.145 (.00302)	−.138 (.00262)	−.144 (.00357)	−.139 (.00257)	−.146 (.00328)	−.147 (.00279)	−.140 (.00220)	−.147 (.00265)	−.147 (.00280)	−.127 (.00378)
β_{11}	.0490 (.0888)	−.0216 (.0129)	−.137 (.00971)	−.268 (.0140)	−.123 (.00778)	−.102 (.00694)	−.317 (.0122)	−.100 (.00637)	−.121	—
β_{12}	.500 (.194)	.139 (.0277)	.142 (.0227)	.170 (.0293)	.0983 (.0219)	—	.147 (.0178)	—	.0671 (.00386)	—
β_{13}	−.555 (.206)	−.0737	.115	−.139	.0790	—	.170	—	.0543	—
β_{M1}	.659 (.864)	.0440 (.0356)	.120	.0407 (.0372)	.0542	−.102	—	−.100	—	—
β_{M2}	3.53 (1.76)	.165 (.113)	.213	.136	.176	−.339	—	−.323	—	—
β_{M3}	−3.94 (1.96)	.265 (.107)	.209 (.104)	.116	.141	−.275 (.0318)	—	−.261	—	—
					NONDURABLES					
α_2	−.467 (.00420)	−.468 (.00353)	−.472 (.00345)	−.464 (.00402)	−.473 (.00385)	−.471 (.00487)	−.461 (.00328)	−.472 (.00440)	−.471 (.00336)	−.495 (.00792)
β_{12}	.272 (.299)	.139 (.0277)	.142 (.0227)	.170 (.0293)	.0983 (.0219)	—	.147 (.0178)	—	.0671 (.00386)	—
β_{22}	−.536 (.774)	−.334 (.0563)	−.306 (.0571)	−.168 (.0619)	−.179	−.339 (.0448)	−.216 (.0326)	−.323	−.241	—
β_{23}	.259 (.866)	.361	.377	.133	.256	—	.0693	—	.174	—
β_{M1}	.282 (.555)	.0440 (.0356)	.120	.0407 (.0372)	.0542	−.102	—	−.100	—	—
β_{M2}	−.271 (1.55)	.165 (.113)	.213	.136	.176	−.339	—	−.323	—	—
β_{M3}	.105 (1.69)	.265 (.107)	.209 (.104)	.116	.141	−.275 (.0318)	—	−.261	—	—

additivity and homotheticity. If we accept additivity, but not homotheticity, we test explicit additivity by imposing one further restriction. If we accept homotheticity, but not additivity, we test homogeneity by imposing one further restriction. The restricted estimates under these restrictions are given in the sixth and seventh columns of Table 1. The corresponding estimates for the indirect translog utility function are given in the sixth and seventh columns of Table 2.

If we accept both additivity and homotheticity hypotheses, we test the hypothesis that the direct utility function is explicitly additive by imposing one further restriction. Similarly, we test homogeneity by imposing one further restriction. Finally, we impose both restrictions under the hypothesis that utility is linear logarithmic. The restricted estimates under each of these sets of restrictions are given in the eighth through tenth columns of Table 1. The corresponding restricted estimates for the indirect translog utility function are given in the eighth through tenth columns of Table 2. The direct and indirect translog utility functions are self-dual if they are linear logarithmic, so that restricted estimates for the two alternative econometric models given in the final columns of the two tables are identical.

C. *Test Statistics*

To test the validity of restrictions implied by the theory of demand and restrictions on the form of the utility function, we employ test statistics based on the likelihood ratio λ, where:

$$\lambda = \frac{\max_{\omega} \mathcal{L}}{\max_{\Omega} \mathcal{L}} \tag{69}$$

TABLE 2—ESTIMATES OF THE PARAMETERS OF THE INDIRECT TRANSLOG UTILITY FUNCTION

Parameters	Unrestricted Estimates	Equality and Symmetry	Additivity	Homotheticity	Additivity and Homotheticity	Explicit Additivity, Not Homotheticity	Homogeneity, Not Additivity	Explicit Additivity and Homotheticity	Additivity and Homogeneity	Linear Logarithmic Utility
					DURABLES					
α_1	−.141 (.00215)	−.125 (.00300)	−.128 (.00279)	−.119 (.00390)	−.119 (.00370)	−.127 (.00228)	−.120 (.00392)	−.120 (.00364)	−.120 (.00370)	−.127 (.00378)
β_{11}	−.115 (.0188)	−.0970 (.0188)	−.0595 (.00972)	−.131 (.0273)	−.0930 (.00823)	−.0662 (.00684)	−.106 (.0246)	−.105 (.00618)	−.0796	—
β_{12}	.695 (.0447)	.0816 (.0300)	.0188 (.0153)	.100 (.0403)	.0282 (.0131)	—	.113 (.0379)	—	.0454 (.00238)	—
β_{13}	−.473 (.0527)	−.0174	.0144	−.0334	.0212	—	−.00754	—	.0342	—
β_{M1}	−.801 (.172)	−.0324 (.0379)	−.0262	−.0644 (.0216)	−.0434	−.0662	—	−.105	—	—
β_{M2}	4.52 (.419)	−.489 (.134)	−.491	−.273	−.183	−.631	—	−.440	—	—
β_{M3}	−3.50 (.464)	−.280 (.114)	−.274 (.107)	−.202	−.137	−.393 (.213)	—	−.330	—	—
					NONDURABLES					
α_2	−.472 (.00472)	−.499 (.00533)	−.494 (.00479)	−.506 (.0576)	−.503 (.00597)	−.497 (.00383)	−.504 (.00612)	−.503 (.00576)	−.502 (.00612)	−.495 (.00792)
β_{21}	−.872 (.120)	.0816 (.0300)	.0188 (.0153)	.100 (.0403)	.0282 (.0131)	—	.113 (.0379)	—	.0454 (.00238)	—
β_{22}	1.11 (.247)	−.668 (.0881)	−.565 (.0715)	−.539 (.0810)	−.300	−.631 (.0307)	−.348 (.0610)	−.440	−.189	—
β_{23}	−.227 (.164)	.0972	.0560	.166	.0891	—	.234	—	.143	—
β_{M1}	−1.63 (.212)	−.0324 (.0379)	−.0262	−.0644 (.0216)	−.0434	−.0662	—	−.105	—	—
β_{M2}	2.98 (.464)	−.489 (.134)	−.491	−.273	−.183	−.631	—	−.440	—	—
β_{M3}	−1.11 (.334)	−.280 (.114)	−.274 (.107)	−.202	−.137	−.393 (.213)	—	−.330	—	—

The likelihood ratio is the ratio of the maximum value of the likelihood function $\mathcal{L}$ for the econometric model of demand ω subject to restriction to the maximum value of the likelihood function for the model Ω without restriction. For normally distributed disturbances the likelihood ratio is equal to the ratio of the determinant of the restricted estimator of the variance-covariance matrix of the disturbances to the determinant of the unrestricted estimator, each raised to the power $-(n/2)$.

Our test statistic for each set of restrictions is based on minus twice the logarithm of the likelihood ratio, or:

$$(70)\quad -2ln\lambda = n(ln\ |\hat{\Sigma}_\omega| - ln\ |\hat{\Sigma}_\Omega|)$$

where $\hat{\Sigma}_\omega$ is the restricted estimator of the variance-covariance matrix and $\hat{\Sigma}_\Omega$ is the unrestricted estimator. Under the null hypothesis this test statistic is distributed, asymptotically, as chi-square with number of degrees of freedom equal to the number of restrictions to be tested.

To control the overall level of significance for each series of tests, direct and indirect, we set the level of significance for each series at .05. We allocate the overall level of significance among the various stages in each series. We first assign a level of significance of .01 to the test of equality and symmetry restrictions implied by the theory of demand. We then assign a level of significance of .04 to tests of restrictions on functional form. These two sets of tests are "nested"; under the null hypothesis the sum of levels of significance provides a close approximation to the level of significance for both sets of tests simultaneously.

We test additivity and homotheticity in parallel, proceeding conditionally on the validity of the theory of demand. These tests are not nested so that the sum of levels of significance for each of the two hypotheses is an upper bound for the level of significance of tests of the two hypotheses considered simultaneously. There are four possible outcomes of our parallel tests of additivity and homotheticity: Reject both, accept both, accept only additivity, and accept only homotheticity. If we reject both hypotheses, our test procedure terminates. If we accept either or both of these hypotheses, we proceed to test explicit additivity or homogeneity, or both.

If we accept both additivity and homotheticity, we test explicit additivity and homogeneity in parallel, conditional on the validity of additivity and homotheticity. Again, the tests are not nested, so that the sum of levels of significance for the two tests provides an upper bound for the tests considered simultaneously. If we accept only additivity, we proceed to test explicit additivity, conditional on additivity. Similarly, if we accept only homotheticity, we proceed to test homogeneity.

Since our three alternative procedures for testing explicit additivity and homogeneity are mutually exclusive, we carry out tests of these hypotheses one time at most. We assign levels of significance to the four hypotheses—additivity, explicit additivity, homotheticity, and homogeneity—with the assurance that the sum of levels of significance provides an upper bound for the level of significance for all four tests. We assign a level of significance of .01 to each of the four hypotheses. With the aid of critical values for our test statistics given in Table 3 the reader can evaluate the results of our tests for alternative allocations of the overall levels of significance among stages of the test procedure.

Table 3—Critical Values of χ^2/Degrees of Freedom

Degrees of Freedom	Level of Significance .10	.05	.025	.01	.005
1	2.71	3.84	5.02	6.63	7.88
2	2.30	3.00	3.69	4.61	5.30
4	1.94	2.37	2.79	3.32	3.72
6	1.77	2.10	2.41	2.80	3.09

Test statistics for both direct and indirect tests of the theory of demand and of restrictions on functional form are presented in Tables 4 and 5. At a level of significance of .01 we reject the hypothesis that restrictions implied by the theory of demand are valid for either direct or indirect series of tests. With this conclusion we can terminate the test sequence, since these results invalidate the theory of demand.

One interesting alternative to our test procedure is to maintain the theory of demand and to test the validity of restrictions on the form of the utility function. Proceeding conditionally on the validity of the theory of demand, we could test the validity of restrictions on the form of the utility function. For the direct series of tests we would reject the restrictions implied by additivity and homotheticity. This conclusion would hold for our preferred procedure of testing these hypotheses in parallel, for a test of either hypothesis conditional on the validity of the

Table 4—Test Statistics for Direct and Indirect Tests of the Theory of Demand and Tests of Additivity and Homotheticity

Hypothesis	Degrees of Freedom	Direct	Indirect
Theory of Demand			
Equality and Symmetry	6	5.64	10.25
Given the Theory of Demand			
Additivity	2	33.63	2.95
Homotheticity	2	13.71	22.68
Additivity and Homotheticity	4	19.02	16.24
Additivity Given Homotheticity	2	24.33	9.81
Homotheticity Given Additivity	2	4.42	29.54

other, or for a joint test of the two hypotheses.

For the indirect series of tests we would reject the restrictions implied by homotheticity, but we would accept the restrictions implied by additivity at a level of significance of .01. For the direct series of tests, we could terminate the test sequence conditional on the theory of demand at this point. For the indirect series of tests, we could test the validity of explicit additivity, given additivity but not homotheticity. We would accept the hypothesis of explicit additivity, given additivity at a level of significance of .01, conditional on the validity of the theory of demand.

In previous econometric studies of demand, the theory of demand has been maintained together with the assumption of additivity. A second alternative to our test procedure is to maintain both the theory of demand and the restrictions implied by additivity. We can test homotheticity; for either direct or indirect series of tests we reject this hypothesis. Given additivity but not homotheticity, we would reject explicit additivity for the direct series of tests and accept explicit additivity for the indirect series of tests. Either of these sets of results would rule out linear logarithmic utility.

TABLE 5—TEST STATISTICS FOR DIRECT AND INDIRECT TESTS OF RESTRICTIONS ON FUNCTIONAL FORM, GIVEN ADDITIVITY OR HOMOTHETICITY OR BOTH

Hypothesis	Degrees of Freedom	Direct	Indirect
	Given Additivity		
Explicit Additivity	1	19.97	1.59
	Given Homotheticity		
Homogeneity	1	1.47	4.73
	Given Additivity and Homotheticity		
Explicit Additivity	1	11.25	4.02
Homogeneity	1	1.74	1.16
Linear Logarithmic Utility	2	29.67	48.31
	Given Explicit Additivity and Homotheticity		
Linear Logarithmic Utility	1	48.09	92.61
	Given Additivity and Homogeneity		
Linear Logarithmic Utility	1	57.60	95.46

IV. Summary and Conclusion

Our objective has been to test the theory of demand without imposing the assumptions of additivity and homotheticity as part of the maintained hypothesis. For either the direct or the indirect series of tests, we conclude that the theory of demand is inconsistent with the evidence. These results confirm the findings of Wold (in association with Lars Juréen) for the double logarithmic demand system and Barten for the Rotterdam demand system.[17] At the same time our results provide the basis for more specific conclusions.

If the theory of demand were valid, the double logarithmic form for the system of demand functions would imply that the utility function is linear logarithmic. Similarly, the validity of the theory of demand and the Rotterdam form for the system of demand functions would imply linear logarithmic utility. An equally valid interpretation of the results of Wold and Barten is that the theory of demand is valid, but that utility is not linear logarithmic. Our results rule out this alternative interpretation and make possible an unambiguous rejection of the theory of demand.

A possible alternative to our test procedure is to maintain the validity of the theory of demand. Proceeding conditionally on the validity of demand theory, we would reject the hypothesis of additivity for the direct series of tests. Additivity of the direct utility function is employed as a maintained hypothesis by Houthakker in the direct addilog demand system, by Stone in the linear expenditure system, and by Basmann, Johansen, and Sato in the demand system incorporating features of both.

---

[17] See Wold, especially pp. 281–302; Barten (1969). A detailed review of tests of the theory of demand is given by Brown and Deaton, pp. 1188–95.

Again, proceeding conditionally on the validity of the theory of demand, we would accept the hypothesis of additivity for the indirect series of tests. This hypothesis is maintained by Houthakker in the implementation of the indirect addilog demand system. Additivity of the direct and indirect utility functions is mutually consistent if and only if the direct utility function is homothetic, so that the acceptance of additivity of the indirect utility function and the rejection of homotheticity rule out acceptance of additivity for the direct utility function.

REFERENCES

G. B. Antonelli, "On the Mathematical Theory of Political Economy," Italy 1886, translated by J. S. Chipman and A. P. Kirman, in Chipman et al., eds., *Preferences, Utility, and Demand,* New York 1971.

A. P. Barten, "Consumer Demand Functions under Conditions of Almost Additive Preferences," *Econometrica,* Jan./Apr. 1964, *32,* 1–38.

———, "Evidence on the Slutsky Conditions for Demand Equations," *Rev. Econ. Statist.,* Feb. 1967, *49,* 77–84.

———, "Maximum Likelihood Estimation of a Complete System of Demand Equations," *Eur. Econ. Rev.,* Fall 1969, *1,* 7–73.

R. L. Basmann, "Hypothesis Formulation in Quantitative Economics: A Contribution to Demand Analysis," in J. Quirk and H. Zarley, eds., *Papers in Quantitative Economics,* Lawrence 1968, 143–202.

———, R. C. Battalio, and J. H. Kagel, "Comment on R. P. Byron's 'The Restricted Aitken Estimation of Sets of Demand Relations,'" *Econometrica,* Mar. 1973, *41,* 365–70.

A. Bergson, "Real Income, Expenditure Proportionality, and Frisch's *New Methods,*" *Rev. Econ. Stud.,* Oct. 1936, *4,* 33–52.

J. A. C. Brown and A. S. Deaton, "Models of Consumer Behaviour: A Survey," *Econ. J.,* Dec. 1972, *82,* 1145–1236.

J. S. Chipman, "Introduction to Part II," in Chipman et al., eds., *Preferences, Utility, and Demand,* New York 1971, 321–31.

L. R. Christensen, D. W. Jorgenson, and L. J. Lau, "Conjugate Duality and the Transcendental Logarithmic Production Function," *Econometrica,* July 1971, *39,* 255–56.

———, ———, and ———, "Transcendental Logarithmic Production Frontiers," *Rev. Econ. Statist.,* Feb. 1973, *55,* 28–45.

J. R. Hicks, "Direct and Indirect Additivity," *Econometrica,* Apr. 1969, *37,* 353–54.

H. Hotelling, "Edgeworth's Taxation Paradox and the Nature of Demand and Supply Functions," *J. Polit. Econ.,* Oct. 1932, *40,* 577–616.

H. S. Houthakker, "Additive Preferences," *Econometrica,* Apr. 1960, *28,* 244–57.

———, "A Note on Self-dual Preferences," *Econometrica,* Oct. 1965, *33,* 797–801.

L. Hurwicz, "On the Problem of Integrability of Demand Functions," in J. S. Chipman, et al., eds., *Preferences, Utility, and Demand,* New York 1971, 174–214.

L. Johansen, "On the Relationships between Some Systems of Demand Functions," *Liiketaloudellinen Aikakauskirja,* no. 1, 1969, 30–41.

L. R. Klein and H. Rubin, "A Constant-Utility Index of the Cost of Living," *Rev. Econ. Stud.,* 1948, *15,* 84–87.

L. J. Lau, (1969a) "Direct and Indirect Utility Functions: Theory and Applications," Inst. Bus. Econ. Res., working pap. no. 149, Berkeley, Mar. 1969.

———, (1969b) "Duality and the Structure of Utility Functions," *J. Econ. Theory,* Dec. 1969, *1,* 374–96.

C. E. V. Leser, "Family Budget Data and Price Elasticities of Demand," *Rev. Econ. Stud.,* Nov. 1941, *9,* 40–57.

D. L. McFadden, "Existence Conditions for Theil-Type Preferences," mimeo., Univ. California, Berkeley 1964.

E. Malinvaud, *Statistical Methods of Econometrics*, 2d ed., Amsterdam 1970.

R. Roy, *De l'Utilité: Contribution à la Théorie des Choix*, Paris 1943.

P. A. Samuelson, "Using Full Duality to Show that Simultaneously Additive Direct and Indirect Utilities Implies Unitary Price Elasticity of Demand," *Econometrica*, Oct. 1965, *33*, 781–96.

———, "Corrected Formulation of Direct and Indirect Additivity," *Econometrica*, Apr. 1969, *37*, 355–59.

K. Sato, "Additive Utility Functions with Double-Log Consumer Demand Functions," *J. Polit. Econ.*, Jan./Feb. 1972, *80*, 102–24.

H. Schultz, *The Theory and Measurement of Demand*, Chicago 1938.

W. H. Somermeyer, J. G. M. Hilhorst, and J. W. W. A. Wit, "A Method for Estimating Price and Income Elasticities from Time Series and its Application of Consumers' Expenditures in the Netherlands, 1949–1959," *Statist. Stud.*, 1961, *13*, 30–53.

J. R. N. Stone, (1954a) *Measurement of Consumers' Expenditures and Behaviour in the United Kingdom*, Vol 1, Cambridge 1954.

———, (1954b) "Linear Expenditure Systems and Demand Analysis: An Application to the Pattern of British Demand," *Econ. J.*, Sept. 1954, *64*, 511–27.

H. Theil, "The Information Approach to Demand Analysis," *Econometrica*, Jan. 1965, *33*, 67–87.

———, *Economics and Information Theory*, Amsterdam 1967.

———, *Principles of Econometrics*, Amsterdam 1971.

H. Wold, *Demand Analysis: A Study in Econometrics*, New York 1953.

Empirical Implications of Infrequent Purchase Behavior in a Stock Adjustment Model

By RICHARD B. WESTIN*

The major component of demand for established consumer durable goods is replacement investment. Various facets of replacement demand have been emphasized in modern studies of consumer durable goods expenditures. The most important consideration has been the necessity of relating depreciation of the service level of the currently held stock to the demand for new stock. This is commonly done in the familiar stock adjustment model of demand by assuming that replacement demand is proportional to net stock holdings.[1] Dale Jorgenson has provided a long-run theoretical basis for this assumption, but its use in empirical studies has been criticized by Martin Feldstein and David Foot for the case of business investment, and by Frank Wykoff (1970, 1973) for the case of automobiles.

Another facet of replacement demand which is often examined is that since service levels are relatively constant over moderate time periods, the timing of replacement of durable goods should be highly discretionary. This view is emphasized by Milton Friedman and has been used in many empirical studies, including recent papers by Saul Hymans and by F. Thomas Juster and Paul Wachtel. This approach is generally considered as an elaboration of the more mechanical stock adjustment approach, however, and the theoretical consistency of combining these models has not been questioned.

This study emphasizes a third facet of replacement demand that is critical to a correct model of the aggregate demand for consumer durables but which has generally been neglected. Individual purchasers of durable goods commonly enter the market only at infrequent intervals and are inactive the rest of the time. This means that the composition of the purchasing public changes in large measure from period to period, and this raises difficult questions of the correct basis for aggregation of the models. For example, the assumption that current replacement demand is equal to current depreciation on the entire existing stock of durables can no longer be justified. Many holders of durables have no intention of entering the market and do not translate the loss of services on their present stock into effective demand for new stock. I shall argue that a correct model of the aggregate demand for durables requires a reinterpretation of the role of substitution in use between new and old durables and an explicit distinction between planned and discretionary replacement.

Although the major question addressed in this paper is technical in nature, the

---

* Scarborough College, University of Toronto. I would like to thank for their helpful comments Ronald Ehrenberg, Robert Ferber, E. Scott Maynes, Paul Wachtel, Susan Westin, Ross Williams, and the referee, none of whom should be held responsible for any errors that remain. Permission of R. L. Polk and Company to use copyrighted data is acknowledged.

[1] For examples of the stock adjustment model applied to automobile demand, see papers by Gregory Chow, Daniel Suits and Gordon Sparks, Hendrik Houthakker and Lester Taylor, Michael Hamburger, and Ross Williams.

policy and forecasting implications of the issues raised are also important. Expenditures for consumer durables are a volatile component of national income, and they have proved to be one of the more difficult sectors to forecast. One implication of the argument raised in this paper is that the stock adjustment models typically used to forecast expenditures on consumer durables are severely biased with respect to the important policy parameters of income and price elasticities of demand. Moreover, the empirical model developed in this paper is directly tested vis-à-vis the stock adjustment model, and the extent of these biases will be demonstrated.

Section I examines the theoretical basis of replacement demand. I shall develop an alternative model of the demand for durable goods that will be conveniently referred to as the discretionary replacement model. Section II considers an empirical test of the discretionary replacement (*DR*) model against the stock adjustment (*SA*) model, and Section III discusses the policy implications of both models. For concreteness, the model is developed in the specific context of the demand for automobiles.

I. Theoretical Model

The theoretical basis for models in which consumers replace their durable goods only at infrequent intervals has been well established. If we assume lump sum transaction costs or imperfect secondhand markets, the optimal purchase policy for individuals is to enter the market only at lengthy intervals and to be inactive otherwise. For purposes of aggregate demand analysis, however, these models are useful only in justifying an approach based on a market where the composition of purchasers changes between periods; a direct derivation of our model from the optimal behavior of individuals would involve difficult assumptions on the distribution of preferences and stocks among individuals. The approach taken is to specify directly an aggregate model while paying particular attention to the empirical implications of infrequent purchase decisions.

To begin our derivation of the *DR* model, we first consider the simplest representation of normal replacement demand. It is assumed that the normal proportion of cars of a given age that will be replaced with new cars in each period is constant.[2] This means that the number of cars replaced with new cars in any year can be adequately predicted as:

$$N_t = \sum_{i=1}^{T} \alpha_i V_{t-i} \tag{1}$$

where N_t is the number of new cars purchased in time t, V_{t-i} is the number of cars of vintage t−i existing in period t, α_i is the proportion of cars of age i that are replaced with new cars, and T is the number of vintages.[3]

Equation (1) does not allow for the possibility of short-run variations in the normal timing of replacement. Consider the case where windfall gains induce some consumers to replace their automobiles one period earlier than they normally would. This would increase replacement demand in the period in which the windfalls occur and decrease replacement demand by the same amount in the next period when the consumers would normally have replaced their cars. To include the effects of short-run variables, the simplest way is to add the current values of these variables to equation (1) to get

---

[2] Replacements must be distinguished from scrappage. New car buyers generally replace cars that are less than five years old, while cars are scrapped much later. The occasionally used assumption that replacement will eventually cause some other car to be scrapped by a process of "bumping" requires an unrealistically rigid view of used car markets.

[3] Stock adjustment models generally use gross expenditures as their dependent variables, but cross-sectional studies (see papers by David Huang (1963, 1964), Maw Lin Lee, De-Min Wu, and John Cragg and Russell Uhler) have shown that the act of purchasing an automobile is determined by different considerations than the amount of gross expenditure on the automobile once the act of purchase has been decided. Our model therefore uses units purchased as the variable to be explained.

$$N_t = \sum_{i=1}^{T} \alpha_i V_{t-i} + \beta D_t \tag{2}$$

where D_t is a discretionary variable. If variations in replacement timing have occurred in previous periods, however, current demand must also reflect these variations. Under the assumption that discretionary variables change the timing of replacement by at most one year in either direction from the consumer's normal replacement date, an adequate procedure for accounting for the previous period's discretionary behavior is to subtract it from current replacement demand to get

$$N_t = \sum_{i=1}^{T} \alpha_i V_{t-i} + \beta(D_t - D_{t-1}) \tag{3}$$

If discretionary variables affect the timing of replacement by more than one period, then a weighted average of past discretionary variables must be subtracted from equation (3).

To understand equation (3), we consider the effects of a windfall gain in period t. The direct effect would be that some consumers who would normally replace their cars in later periods would replace them in period t, earlier than they had planned, and we get equation (2). In period t+1, however, assuming no further windfall gains, equation (2) would predict a return to normal replacement levels when in actuality some of the normal replacement that should have occurred in period t+1 has already occurred in period t. We would therefore expect a fall in purchases in period t+1 below normal replacement levels, and this is accounted for in equation (3). In period t+2, if no further windfall gains or losses occur, the market would return to normal replacement demand.

Replacement demand is important, but it is not a sufficient explanation of total demand. We must also account for nonreplacement demand which encompasses new entrants into the automobile market, increasing multiple car ownership among existing owners, and families that "trade up" to become new car buyers. Nonreplacement demand is assumed to be determined by observable economic variables such as income and relative prices, and by the existing stock of cars which represents a saturation effect on the market for new owners. Preserving the linear form of our specification gives us a final demand equation of the form:

$$N_t = N_t^* + \sum_{i=1}^{T} (\alpha_i - \gamma_i) V_{t-i} + \beta(D_t - D_{t-1}) \tag{4}$$

where N_t^*, nonreplacement demand, is a function of observable economic variables, and γ_i represents the saturation effect of the existing stock on nonreplacement demand.

The discretionary replacement model has been deliberately constructed in such a way that it is directly testable against the traditional stock adjustment model. Since contrasting these two models will clarify the theoretical role of a market where the composition of purchasers changes between periods, we turn briefly to the *SA* model before proceeding. The familiar *SA* model is expressed as:

$$A_t = \theta(S_t^* - S_{t-1}) + R_t \tag{5}$$

where A_t is new purchases in period t measured either as gross investment or as the number of units purchased; S_t^* consumers' desired net stock is a function of observable economic variables; S_{t-1} is existing undepreciated stock at the end of period t−1, measured either as net stock or the number of units in existence; R_t replacement demand in period t makes good the depreciation of the net stock; and θ is a constant between zero and one. Under the assumption that depreciation occurs at a constant rate $(1-k)$, equation (5) can be rewritten as

$$(6) \quad A_t = \theta S_t^* + (1 - k - \theta) \sum_{i=0}^{\infty} k^i A_{t-i-1}$$

Comparing the *DR* model (equation (4)) with the *SA* model (equation (6)), we note that aside from minor differences in the dependent variables and the interpretations of the coefficients, the two equations are equivalent except for the addition of the term $\beta(D_t - D_{t-1})$ to equation (4). As current discretionary variables have also been included in traditional *SA* models of new car demand, this means that the unique feature of the *DR* model is the inclusion of lagged discretionary variables which are specified to have an effect opposite in sign and equal in magnitude to current discretionary variables.[4] Since these variables are at the core of our criticism of *SA* models, it is necessary to contrast the assumptions of both models to justify our empirical tests.

The *SA* model is generally justified by an implicit model of continuous adjustment behavior in which the existing stock of durables has both a positive and negative effect on the demand for new durables—positive because of the necessity to replace the existing stock as it wears out and negative because it provides a substitute in use for new durables. Although the *SA* and *DR* models differ in measuring the first of these effects (i.e., whether to use depreciation or replacement patterns), the crucial difference between the models is in the concept of substitution in use between existing stocks and new durables, which is generally known as the inventory effect.

The inventory effect of existing stock on purchase rates is indeed powerful. The important distinction, however, is that its role is not to slow the rate of net investment in a continuous adjustment model but rather to preclude many consumers from entering the market at all in any given time period. More concretely, if car owners who own very new cars will not enter the market again for a long period of time, their holdings of stock are irrelevant for a prediction of demand in the near future since these consumers are effectively removed from the market. It is at this point that we encounter the distributional effects of a market where the composition of purchasers changes from period to period. The *SA* model specifies that very new used cars have a powerful depressing effect on current expenditures because they provide a substitute in use, but the fact that potential new car purchasers are not the same people as those who recently bought cars means these substitution possibilities do not exist. In contrast, the *DR* model we have proposed says that because we shift to a new group of potential purchasers, the stock of very new used cars is irrelevant for an explanation of demand, save for the relatively minor effects of some normal replacement demand and the saturation effect on nonreplacement demand.

Extensive empirical experience with the *SA* model indicates that equation (6), in which the existing stock has a negative effect on demand that decreases in absolute magnitude with age, provides a good explanation of demand. In the *DR* model, the saturation effect of the existing stock on nonreplacement demand makes the signs of the coefficients of existing stock indeterminant, but qualitatively our emphasis on replacement demand implies that the coefficients of existing stock might be expected to be positive or close to zero. Since this hypothesis on the coefficients is in contradiction to the known empirical results of the *SA* model, it is necessary to give a theoretical reason why existing stock has been consistently found to have a strong negative influence.

---

[4] Juster and Wachtel derive a stock adjustment model in which both current and lagged discretionary variables enter as separate variables with coefficients specified to be opposite in sign. The occurrence of the lagged discretionary variables in their equation is a result of the reduction of their structural model to a reduced form before estimation, however, and the coefficients of the lagged discretionary variables are overidentified in terms of the structural parameters of the stock adjustment model.

As we emphasize, the major difference between the models is that the *SA* model omits the lagged values of discretionary variables, which in our model represent the effects of past variations in planned replacement behavior. If the *DR* model correctly describes aggregate demand, the effect of omitting the lagged discretionary variables in the *SA* model is to bias the coefficients of the other variables, particularly those variables most highly correlated with the omitted variables. Since one-period lagged discretionary variables are also a determinant of one-period old stock, we would expect the newest existing stock variables to exhibit the maximum correlation with the lagged discretionary variables, and hence the coefficients of the newest existing stock variables should suffer the maximum bias when lagged discretionary variables are omitted from the the equation. Furthermore, since the effect of lagged discretionary variables on current purchases is postulated to be opposite in sign to their previous period's effect on purchases, the bias in the coefficient of one-period lagged stock would be negative. More concretely, our model postulates that windfall gains would have a positive effect on purchases in the current period and a negative effect in the next period. If one-period lagged windfall gains are omitted from the equation, their effect would be represented by a negative coefficient on one-period lagged stock. We therefore argue that the consistent finding of negative stock coefficients for the *SA* model may not be due to the inventory effect of existing stock on new purchases, but rather is due to the fact that existing stock variables are serving as proxy variables for omitted lagged discretionary variables.

These last considerations suggest an empirical test of the two proposed models. Instead of entering discretionary variables as first differences in equation (4), they can be entered independently and their separate contribution estimated. If the *SA* model is correct in that automobile demand is dominated by the inventory effect of existing stock, we would expect the coefficients of lagged discretionary variables to be insignificantly different from zero and the coefficients of existing stock variables to be negative and decreasing in absolute magnitude for older cars. On the other hand, if the correct model is the *DR* model, which emphasizes replacement demand and the effects of short-run variations in purchase timing, then the coefficients of lagged discretionary variables should be opposite in sign and roughly equal in magnitude to the coefficients of current discretionary variables and the coefficients of existing stock variables should be positive or close to zero. We turn now to our empirical results.

II. Empirical Results

The models specified in the previous section were estimated using annual data for the period 1953–72, with data from 1951 and 1952 used to give starting values of the lagged discretionary variables. The dependent variable was total new passsenger car retail sales.[5]

---

[5] These data unfortunately include business purchases of passenger cars. The alternative of using deflated expenditures data from the *Survey of Current*

The discretionary variables used in this study were the unemployment rate, the Survey Research Center Index of Consumer Sentiment, and transitory income; these variables have an extensive history of use in automobile demand studies (see, for example, Irwin Friend and F. Gerald Adams, Hymans, and Juster and Wachtel.)[6,7] The variables used to explain nonreplacement demand were disposable and permanent income and the relative price of new automobiles. The coefficients of the lagged stock variables in the regressions reported here were constrained to lie on a third-degree polynomial with the side restriction that the polynomial equal zero after seven periods. All equations have been reestimated under alternative specifications on the length and degree of the polynomial restriction, including the direct entry of the lagged stock variables; none of the substantive conclusions of this paper was affected by the use of the polynomial restriction.[8] Finally, a dummy variable representing strikes was added to all equations.

Table 1 summarizes the regression results for the models discussed in the last section. In order to judge the contribution of the lagged stock variables to the estimated equations, the F-statistic is used to test the hypothesis that the coefficients of the lagged stock variables all equal zero. The values of the F-statistic are reported in Table 1 under the heading "partial F." In addition, a key test of the *DR* model is the hypothesis that the coefficient of a lagged discretionary variable should be opposite in sign and equal in magnitude to the coefficient of the current value of that discretionary variable; the values of the t-statistic used to test this hypothesis are presented in Table 2.

Results 1, 2, and 3 in Table 1 correspond to the *SA* model, and they are compatible with the a priori predictions of that model. In all three results, the lagged stock variables make a significant contribution to the explained variation, and their estimated coefficients are large and negative and (once current discretionary variables are included) decline in absolute value for older stock.

Results 4–11 in Table 1 provide a test of the *DR* model we have proposed. Result

---

Business does not solve this problem, as the adjustment made for business expenditures in the *OBE* data is to reduce gross expenditures by a constant 17 percent (see Lawrence Grose and Edward Bassett).

[6] Current and lagged values of first differences of the discretionary variables were also tested to see whether effects of discretionary variables were best measured as deviations from long-run averages (implying the use of gross values of the variables) or as deviations from immediate past values (implying the use of the first differences); the former alternative fit the data better. Weighted averages of one- and two-period lagged discretionary variables were tried in order to investigate the hypothesis that short-run variations in replacement behavior affect purchases over a period of more than one year; the attitudes variable was improved by putting more weight on two-period lagged values, but these results are not reported here. I also tested the 4–6 month prime commercial paper rate as a representation of credit conditions but it performed poorly, probably because it represented a nominal rather than a real interest rate. Finally, lagged income and price variables were included to test the extent to which they contain discretionary influences, but the results were insignificant in the presence of other variables.

[7] Papers by Chow, Houthakker and John Haldi, and Wu argue that since the regression coefficient of current transitory income is not significantly different from the regression coefficient of permanent income in their studies, the division of disposable income into permanent and transitory components is irrelevant. This conclusion misses the essential point of Friedman's argument, which is that transitory income affects purchase timing while permanent income affects gross investment. The key issue is therefore not whether current transitory income has the same impact on purchases as permanent income, but whether *lagged* transitory income has an equal and opposite effect to current transitory income on purchases, an effect postulated by the *DR* model.

[8] It can be argued that this polynomial restriction forces any bias that occurs in the coefficient of one-period lagged stock to be transmitted to the coefficients of the other lagged stock variables. This effect occurs in the *SA* model when depreciation rates or scrappage rates are used to construct the stock variable, so a polynomial restriction on the coefficients is a compromise between the rigid weights of the *SA* model and the flexible pattern of the *DR* model.

TABLE 1—REGRESSION ESTIMATES OF STOCK ADJUSTMENT AND DISCRETIONARY REPLACEMENT MODELS[a]

	Results										
	1	2	3	4	5	6	7	8	9	10	11
Intercept	11.7	9.8	9.6	18.5	9.0	14.6	12.3	19.1	12.8	15.6	12.8
YD	.033	.025		.011	.018	.015	.018				
	(3.88)	(3.84)		(1.46)	(3.80)	(1.59)	(3.36)				
YP			.025					.011	.010	.011	.015
			(3.56)					(0.79)	(0.78)	(0.69)	(2.47)
PCR	−8.2	−12.3	−12.0	−16.1	−12.3	−14.3	−9.4	−16.7	−13.5	−16.1	−10.0
	(−1.48)	(−2.94)	(−2.51)	(−3.64)	(−2.85)	(−1.82)	(−2.79)	(−2.91)	(−2.35)	(−1.67)	(−2.84)
UE		−.144	−.140	−.303		−.239		−.314		−.262	
		(−1.23)	(−1.07)	(−3.10)		(−1.71)		(−2.28)		(−1.55)	
UE_{-1}				.591		.363		.619		.397	
				(4.00)		(1.38)		(2.91)		(1.23)	
ΔUE							−.321				−.349
							(−3.76)				(−3.64)
ATT		.070	.068		.083	.035			.090	.039	
		(3.58)	(2.68)		(4.88)	(1.12)			(3.09)	(1.05)	
ATT_{-1}					−.018	−.008			−.012	−.005	
					(−0.72)	(−0.23)			(−0.46)	(−0.13)	
ΔATT							.030				.031
							(1.83)				(1.77)
YT			.032					.003	.038	.001	
			(0.81)					(0.06)	(0.88)	(0.01)	
YT_{-1}								.004	−.049	−.015	
								(0.09)	(−0.84)	(−0.26)	
ΔYT											.010
											(0.32)
V_{-1}	−.18	−.35	−.37	.04	−.14	−.08	.03	.05	−.04	.01	.10
V_{-2}	−.25	−.20	−.21	.05	−.07	−.02	−.03	.07	.08	.08	.03
V_{-3}	−.28	−.11	−.12	.01	−.04	−.02	−.08	.02	.10	.06	−.04
V_{-4}	−.28	−.08	−.08	−.06	−.05	−.06	−.12	−.06	.06	−.01	−.10
V_{-5}	−.25	−.08	−.06	−.12	−.07	−.10	−.14	−.13	−.01	−.10	−.14
V_{-6}	−.19	−.07	−.06	−.16	−.09	−.12	−.13	−.17	−.06	−.15	−.14
V_{-7}	−.10	−.06	−.04	−.12	−.07	−.10	−.08	−.13	−.07	−.13	−.10
$STRIKE$	.402	.296	.299	.516	.335	.440	.476	.513	.444	.479	.512
	(1.91)	(2.25)	(2.15)	(4.01)	(2.08)	(2.80)	(4.00)	(2.55)	(2.03)	(2.21)	(3.56)
R^2	.930	.977	.977	.980		.983	.981	.980		.983	.981
$S.E.$	.577	.355	.374	.333	.364	.345	.322	.370	.390	.390	.341
Partial F	2.67^b	3.72^c	3.27^b	2.48	1.67	1.52	2.34	1.37	0.66	1.46	1.88
^					−.30				−.40		

Note: The independent variables are as follows:

YD = disposable income, 1958 dollars
YP = permanent income, 1958 dollars
PCR = relative price of new cars
UE, UE_{-1}, ΔUE = unemployment rate (current, lagged, first difference)
ATT, ATT_{-1}, ΔATT = Index of Consumer Sentiment (current, lagged, first difference)
YT, YT_{-1}, ΔYT = transitory income (current, lagged, first difference)
V_{-i} = number of cars in use on January 1 of vintage i years old
$STRIKE$ = dummy variable for strikes

[a] The figures in parentheses are t-ratios; $\hat{\rho}$ is the estimated autocorrelation coefficient.
[b] Denotes significance at the 10 percent level.
[c] Denotes significance at the 5 percent level.

4 provides strong confirmation of this model; the regression coefficient of the lagged unemployment rate is opposite in sign to the coefficient of the current value of the unemployment rate, and the test statistics show that we can strongly reject the null hypothesis that the coefficient of the lagged unemployment rate is equal to zero while we cannot reject (by Table 2) the hypothesis that the coefficient is equal in magnitude and opposite in sign to the coefficient of the current unemployment rate. Furthermore, the coefficients of the lagged stock variables are positive for the newest stock, and the partial F-test indicates that the stock variables in total do not make a significant contribution to the explained variation of this equation. These considerations on the coefficients of the stock variables are important because the sign and magnitude of the coefficient of the lagged unemployment rate could be

justified in a *SA* model in which the first difference in the unemployment rate determined desired stock. In this case, however, the coefficients of the lagged stock variables should still be negative and decline in absolute value as the stock gets older, so the coefficients of the lagged stock variables in result 4 of Table 1 indicate this hypothesis is not valid.

Although entering the unemployment rate as a discretionary variable produces results which are strongly compatible with the *DR* model, when the Index of Consumer Sentiment was entered as the only discretionary variable the results favored the *SA* model. The residuals of the equation showed negative autocorrelation was present, however, and because discretionary variables are postulated to enter with both current and lagged values and with coefficients opposite in sign in the *DR* model, negative autocorrelation indicates an alternative discretionary variable was omitted. We therefore reestimated the equation by maximum likelihood methods assuming a first-order Markov process for the error terms.[9] The results are given as 5 in Table 1 and favor neither model strongly. Rather than pursue the question of autocorrelation further, however, it is more important to identify the omitted discretionary variables, and we turn to the next equation.

Result 6 in Table 1, where current and lagged values of both the unemployment rate and the Index of Consumer Sentiment are entered as separate variables, is a severe drain on the available degrees of freedom, and the test statistics on the coefficients are inconclusive. The magnitudes of the coefficients, however, are most compatible with the *DR* model. We then entered the discretionary variables as first differences in result 7. This gives the formulation originally derived as the specification of the *DR* model and requires the same number of coefficient estimates as the *SA* model with current discretionary variables included; the models can then be compared by examining results 2 and 7 in Table 1. We find that entering the discretionary variables as first differences gives coefficients of the lagged stock variables that are small and positive for the newest stock; these results are consistent with the *DR* model and inconsistent with the *SA* model. Furthermore, by examining the goodness of fit statistics, R^2, and the standard error of the equation, we see that the *DR* model has a slightly better fit to the data than the *SA* model. The small difference in goodness of fit between the two models is not surprising. Our earlier discussion of the possible bias in the *SA* model emphasized that the newest lagged stock variables would be efficient proxies if the lagged discretionary variables were omitted. The important point is that the *DR* model achieves as good a fit as the traditional *SA* model while implying that the powerful inventory effect postulated by the *SA* model is

TABLE 2—TEST STATISTICS ON SUMS OF COEFFICIENTS[a]

Regression Result	Discretionary Variables		
	UE and UE_{-1}	ATT and ATT_{-1}	YT and YT_{-1}
4	1.57		
5		2.58**	
6	0.41	0.67	
8	1.41		0.11
9		2.18*	−0.17
10	0.41	0.72	−0.20

[a] t-statistics for the null hypothesis that the sum of the regression coefficients for the current and the lagged values of a discretionary variable equals zero. The symbol * denotes the hypothesis is rejected at the 10 percent significance level; the symbol ** denotes rejection at the 5 percent significance level.

---

[9] The test statistics of results 5 and 9 in Tables 1 and 2 are overstated as they do not correct for the fact that the autocorrelation parameter has been estimated (see J. Phillip Cooper).

insignificant. We have therefore found that the a priori predictions of the *DR* model are borne out by the data.

The rest of the results in Table 1, 8 through 11, repeat those discussed above with a measure of permanent income substituted for disposable income and transitory income treated as a discretionary variable. The results are compatible with their counterpart equations discussed above and we will not discuss them in depth other than to observe that again the results are more compatible with the *DR* model than the *SA* model.

Before turning to the policy implications of the *DR* model, we note that the bias we have been discussing would also exist for other versions of the *SA* model that eliminate most of the lagged stock variables from the equations to be estimated. Under the assumption that depreciation on the existing stock occurs at a constant rate, Marc Nerlove and Houthakker and Taylor show that the *SA* model can be transformed to eliminate all existing stock variables other than one-period lagged purchases. As should be clear from our discussion, however, the bias from omitting lagged discretionary variables centers mainly on one-period old stock, so the results of these models would still be biased. To test this claim, we estimated *SA* models in the Nerlove and Houthakker and Taylor forms with current and lagged discretionary variables included. Rather than examine these results in detail, it is sufficient to examine the estimated depreciation rates implied by the models. If the *SA* model is correct, the estimated depreciation rate should lie in the range of 20 to 30 percent on an annual basis, which contains the rates estimated from used car price data (see Chow and Wykoff (1970)). On the other hand, if the *DR* model correctly explains purchase behavior, one can argue that the estimated depreciation rate would be higher when lagged discretionary variables are included. Referring to the *SA* model, equation (6), we see that small values of k (which equals one minus the depreciation rate) would imply small coefficients attached to the lagged stock variables, which is predicted by the *DR* model. Since the Nerlove and Houthakker and Taylor models are basically a transformation of equation (6) where k is estimated, one would therefore expect that adding current and lagged discretionary variables to equation (6) and applying the transformation would yield a small value of k and hence a high depreciation rate if the *DR* model is correct.

When the *SA* model in the Nerlove form was estimated on our data with current discretionary variables included, the estimated depreciation rate ranged from 5 to 30 percent on an annual basis, which is compatible with the predictions from used car prices.[10] When lagged discretionary variables were included, the estimated depreciation rate jumped to a range of 45 to 70 percent on an annual basis. Since this is what one would expect when the inclusion of lagged discretionary variables removes the downward bias in the estimated effect of lagged stock variables, it is clear that the Nerlove and Houthakker and Taylor versions of the *SA* model do not escape the bias discussed here. We turn now, therefore, to the policy implications of our results.

III. Policy Implications

Up to this point, we have emphasized the bias in the regression coefficients of existing stock variables in the *SA* model caused by a neglect of short-run discretionary replacement activity. From an economic policy standpoint, however, the major interest in estimating demand equations would be on the effect of income and prices, as existing stock variables are not subject to exogenous control. If we examine closely the two models considered in the preceding section, we find that the biases are not limited solely to the coefficients of the stock variables but also affect our estimates of other key parameters. In particular, we will show that estimated demand elasticities with respect to income and prices are sensitive to the model specification we choose.

---

[10] The attitudes variable was omitted from these tests as it predicted an absurdly high depreciation rate for the *SA* model. Results of tests on the Houthakker and Taylor model were similar to that of the Nerlove model.

Discretionary purchase behavior causes short-run variations in purchases, but automobile purchases have also shown a long-term upward trend because of new entrants into the automobile market. Of the independent variables we have used in our empirical analysis, the discretionary variables have shown no strong trend in this period, nor has the relative price of new cars; the independent variables that do show a trend are the two income measures and the lagged stock variables. As the lagged stock variables are postulated to have either a negative effect (*SA* model) or a small effect (*DR* model) on purchases, the brunt of the explanation of the upward trend in automobile purchases must be incorporated into the regression coefficient of disposable or permanent income. The overall growth trend implied by the equation is then determined by the extent to which the upward trend in purchases caused by the growth of income is offset by the effects of growth of existing automobile stocks. This overall growth trend can be summarized by examining the long-run elasticity of purchases with respect to income, which is computed as

$$\left[b \Big/ \left(1 - \sum_{i=1}^{T} a_i\right)\right] \frac{\text{income}}{\text{purchases}} \tag{7}$$

where b is the estimated regression coefficient of disposable or permanent income and $\sum_{i=1}^{T} a_i$ is the sum of the estimated regression coefficients of the lagged stock variables (V_{-i}).

TABLE 3—AVERAGE PURCHASE ELASTICITIES[a]

	Income Elasticities		Price Elasticities	
	Long-run	Short-run	Long-run	Short-run
Stock Adjustment Model[b]	0.68	1.46	0.64	1.32
Discretionary Replacement Model[c]	0.52	0.72	1.26	1.65

[a] Elasticities are evaluated at the sample means of the variables.
[b] Results 1–3, Table 1.
[c] Results 4–11, Table 1.

Since the actual growth trend in purchases is implicit in the data, we would expect all the results reported in Table 1 to show roughly the same long-run income elasticity. This can be verified by examining the average purchase elasticities reported in Table 3; the estimated long-run elasticity of new car purchases with respect to income is similar for both the *DR* and *SA* models. On the other hand, the short-run elasticity of new car purchases with respect to income is very much higher for the *SA* model than the *DR* model. This elasticity is defined as $b \cdot (\text{income}/\text{purchases})$ and depends only on the regression coefficient of income. Now as the estimated coefficients of the lagged stock variables are biased downwards in the *SA* model, equation (7) shows that the regression coefficient of income must be biased upwards to preserve the long-term growth trend implied by the equation.[11] This means that estimates of short-run income elasticities, an important policy parameter in short-run stabilization questions, will be biased upwards in the *SA* model.

---

[11] More formally, if a variable is omitted from a linear regression, the bias in the coefficients of the included variables is proportional to the coefficients that would be obtained in a linear regression of the omitted variables on the included variables. Regressions we performed of the lagged discretionary variables on the other explanatory variables verified the strength and direction of the multiple biases discussed in this section of the text.

Turning to the long- and short-run price elasticities, we would not expect the regression coefficient of relative price to be seriously affected by a change from a *SA* model to a *DR* model because there is no clear correlation between prices and lagged discretionary variables. Because the coefficients of the lagged stock variables are biased downward in a *SA* model, however, we would expect the corresponding long-run price elasticity to be understated. This is what we observe in Table 3, where the average long-run price elasticity for the *SA* model is significantly below that of the *DR* model. Indeed, we observe that both the long- and short-run price elasticities for the *DR* model are above unity, which contradicts the usual finding of the *SA* model.

IV. Summary

This paper has argued that the appropriate theoretical specification for aggregate demand models for consumer durables must be a discrete purchase model rather than a continuous adjustment model. The key issue is the concept of substitution in use between new and old durables. In order to assume that very new used stock depresses new purchases because it is a good substitute in use requires the strong assumption that the distribution of stock among individuals is irrelevant. Alternatively, we can realize that recent purchasers are not the same people as those currently entering the market, and the stock holdings of recent purchasers are irrelevant for predicting current demand. Our empirical emphasis on discretionary variables in this paper has then come about because of the statistical problems of distinguishing discretionary replacement behavior from the inventory effect of existing stocks.

The discretionary replacement model proposed here requires data on existing stocks by vintages. Except for automobiles, these data are rarely available except as an extrapolation of purchase data. Our empirical work has found, however, that in terms of explaining the statistical variation of purchases over time, the data on existing stocks are unimportant. This is not surprising; the role of stock in the discretionary replacement model is basically an echo effect, and the echo effect is weak, even for automobiles. The surprising implication for forecasting is that existing stocks are relatively unimportant statistical variables, even though their theoretical role in the model is crucial.

The stock adjustment model has been the principal model for analyzing the demand for consumer durables for twenty years. Its use, however, is based on the unjustified assumption that aggregation lets the discrete purchase behavior of individuals be represented as a continuous adjustment process. This paper shows that this approach severely biases the estimated equations. It proposes an alternative model where the implications of infrequent purchase behavior can be explicitly treated.

DATA APPENDIX

Sources: For data listed as taken from the *Survey of Current Business*, data for years prior to 1971 were taken from the 1971 edition of *Business Statistics*, the biennial supplement to the *Surv. Curr. Bus.* Data for 1971 and 1972 were taken from the July 1973 issue of the *Surv. Curr. Bus.*

Total new passenger car retail sales: For 1966–72, *Surv. Curr. Bus.*; prior to 1966, redefined as new passenger car registrations, *Surv. Curr. Bus.*

Disposable income: Surv. Curr. Bus., various issues.

Permanent income: Defined using weights given in M. Friedman, p. 147, plus an estimated time trend. The time trend was introduced because otherwise transitory income was positive for early and late years and negative for middle years.

Transitory income: Defined as real disposable income minus permanent income.

Relative price of new cars: Defined as the ratio of the implicit price deflator for new autos and parts divided by the implicit price deflator for personal consumption expenditures, *Surv. Curr. Bus.* The *BLS* price series for new cars, *Surv. Curr. Bus.*, was also tried and rejected.

Unemployment rate: Surv. Curr. Bus., various issues.

Index of Consumer Sentiment: For 1969–72, data from the 1970–73 annual issues of the *Survey of Consumer Finances*; for years prior to 1969, the revised Index was obtained directly from the Survey Research Center, University of Michigan.

Number of cars in use on January 1 by year of production: For years prior to 1965, data from paper by Charles Friedman; for 1965–69, unpublished data obtained from the Office of Business Economics, U.S. Department of Commerce; for 1970–72, data constructed by extrapolation using scrappage rates given in paper by Franklin Walker.

Strike: Since automobile strikes typically occur near the end of the year and cause postponement of purchases until the beginning of the next year, the strike dummy variable was defined as −1 in 1964, 1967, and 1970 (auto strikes) and in 1959 (steel strike) and as +1 in years following the strike years.

REFERENCES

G. Chow, "Statistical Demand Functions for Automobiles and Their Use for Forecasting," in A. Harberger, ed., *The Demand for Durable Goods*, Chicago 1960, 149–78.

J. P. Cooper, "Asymptotic Covariance Matrix of Procedures for Linear Regression in the Presence of First Order Serially Correlated Disturbances," *Econometrica*, Mar. 1972, *40*, 305–10.

J. Cragg and R. Uhler, "The Demand for Automobiles," *Can. J. Econ.*, Aug. 1970, *3*, 386–406.

M. Feldstein and D. Foot, "The Other Half of Gross Investment: Replacement and Modernization Expenditures," *Rev. Econ. Statist.*, Feb. 1971, *53*, 49–58.

C. Friedman, "The Stock of Automobiles in the United States," *Surv. Curr. Bus.*, Oct. 1965, *45*, 21–27.

M. Friedman, *A Theory of the Consumption Function*, Princeton 1957.

I. Friend and F. G. Adams, "The Predictive Ability of Consumer Attitudes, Stock Prices, and Non-Attitudinal Variables," *J. Amer. Statist. Assn.*, Dec. 1964, *59*, 988–1005.

L. Grose and E. Bassett, "Automobile Purchases by Business and Consumers as Reflected in the National Accounts," *Surv. Curr. Bus.*, Mar. 1962, *42*, 13–17.

M. Hamburger, "Interest Rates and the Demand for Consumer Durable Goods," *Amer. Econ. Rev.*, Dec. 1967, *57*, 1131–53.

H. Houthakker and J. Haldi, "Household Investments in Automobiles: An Intertemporal Cross-Section Analysis," in I. Friend and R. Jones, eds., *Consumption and Saving*, Vol. 1, Philadelphia 1960, 175–224.

——— and L. Taylor, *Consumer Demand in the United States*, 2d ed., Cambridge 1970.

D. Huang, "Initial Stock and Consumer Investment in Automobiles," *J. Amer. Statist. Assn.*, Sept. 1963, *58*, 789–98.

———, "Discrete Stock Adjustment: The Case of Demand for Automobiles," *Int. Econ. Rev.*, Jan. 1964, *5*, 46–62.

S. Hymans, "Consumer Durable Spending: Explanation and Prediction," *Brookings Papers*, Washington 1970, *2*, 173–206.

D. Jorgenson, "Anticipations and Investment Behavior," in J. Duesenberry et al., eds., *The Brookings Quarterly Econometric Model of the United States*, Chicago 1965, 35–94.

F. T. Juster and P. Wachtel, "Anticipatory and Objective Models of Durable Goods Demand," *Amer. Econ. Rev.*, Sept. 1972, *62*, 564–79.

M. L. Lee, "Income, Income Change, and Durable Goods Demand," *J. Amer. Statist. Assn.*, Dec. 1964, *59*, 1194–1202.

M. Nerlove, "The Market Demand for Durable Goods: A Comment," *Econometrica*, Jan. 1960, *28*, 132–42.

D. Suits and G. Sparks, "Consumption Regressions with Quarterly Data," in J. Duesenberry et al., eds., *The Brookings Quarterly Econometric Model of the United States*, Chicago 1965, 203–23.

F. Walker, "Determinants of Auto Scrappage," *Rev. Econ. Statist.*, Nov. 1968, *50*, 503–06.

R. Williams, "Demand for Consumer Durables: Stock Adjustment Models and Alternative Specifications of Stock Depletion," *Rev. Econ. Stud.*, July 1972, *39*, 281–96.

D. Wu, "An Empirical Analysis of Household Durable Goods Expenditures," *Econometrica*, Oct. 1965, *33*, 761–80.

F. Wykoff, "Capital Depreciation in the Postwar Period: Automobiles," *Rev. Econ. Statist.*, May 1970, *52*, 168–72.

———, "A User Cost Approach to New Automobile Purchases," *Rev. Econ. Stud.*, July 1973, *40*, 377–90.

Survey Research Center, *Survey of Consumer Finances*, Ann Arbor, 1970–73.

U.S. Office of Business Economics, *Surv. Curr. Bus.*, various issues, Washington 1966–73.

———, *1971 Business Statistics*, *Surv. Curr. Bus.* suppl., Washington 1972.

The Deterrent Effect of Capital Punishment: A Question of Life and Death

By ISAAC EHRLICH*

Debate over the justness and efficacy of capital punishment may be almost as old as the death penalty itself. Not surprisingly, and as is generally recognized by contemporary writers on this topic, the philosophical and moral arguments for and against the death penalty have remained remarkably unchanged over time (see Thorsten Sellin (1959, p. 17), and (H. A. Bedau, pp. 120–214). Due in part to its essentially objective nature, one outstanding issue has, however, become the subject of increased attention in recent years and has played a central role in shaping the case against the death penalty. That issue is the deterrent effect of capital punishment, a reexamination of which, in both theory and practice, is the object of this paper.

The multifaceted opposition to capital punishment relies partly upon ethical and aesthetic considerations. It arises also from recognition of the risks of errors of justice inherent in a legal system, errors occasionally aggravated by political, cultural, and personal corruption under certain social regimes. Such errors, of course, are irreversible upon application of this form of punishment. But the question of deterrence is separable from subjective preferences among alternative penal modes and can be studied independently of any such preferences. Of course, the verification or estimation of the magnitude of the deterrent effect of the death penalty—the determination of the expected tradeoff between the execution of a murderer and the lives of potential victims it may help save—can, in turn, influence evaluation of its overall desirability as a social instrument even if that evaluation is largely subjective.

Recent applications of economic theory have presented some analytical considerations and empirical evidence that support the notion that offenders respond to incentives and, in particular, that punishment and law enforcement deter the commission of specific crimes. Curiously, two of the most effective opponents of capital punishment, Beccaria in the 18th century and Sellin in recent years, have never to my knowledge questioned analytically the validity of the deterrent effect of punishment in general. Beccaria even recognizes explicitly the probable existence of such a general effect. What has been questioned by these scholars is the existence of a *differential* deterrent effect of the death penalty over and above its most common practical alternative, life imprisonment (see Beccaria, pp. 115–17). Sellin has presented extensive statistical data that he and others have interpreted to imply, by and large, the absence of such an effect (see Sellin (1959, 1967)).

Whether, in fact, the death penalty constitutes a more severe punishment than life imprisonment for the average potential offender cannot be settled on purely logical grounds, although crime control legislation, ancient and modern, clearly answers this question affirmatively. Observation that convicted offenders almost universally seek and welcome the commutation of a death sentence to life imprisonment is consistent with the intuitive ranking of the death penalty as the harshest of all punishments. Still, one may argue that the differential deterrent effect of capital punishment on the incentive to commit murder may be offset by the added incentive it may create for those who actually commit this crime to eliminate policemen and witnesses who can bring about their apprehension and subsequent conviction and execution.

---

* University of Chicago and National Bureau of Economic Research. I have benefitted from comments and suggestions from Gary Becker, Harold Demsetz, Lawrence Fisher, John Gould, Richard Posner, George Stigler, and Arnold Zellner. I am particularly indebted to Randall Mark for useful assistance and suggestions and to Walter Vandaele and Dan Galai for helpful computational assistance and suggestions. This paper is a reduced version of a more complete and detailed draft (see the author 1973b). Financial support for this study was provided by a grant to the NBER from the National Science Foundation, but the paper is not an official NBER publication since it has not been reviewed by the board of directors.

The existence of the differential deterrent effect of capital punishment is ultimately an empirical matter. It cannot, however, be studied effectively without thorough consideration of related theoretical issues. The crucial empirical question concerns the kind of statistical test to devise in order to accept or reject the relevant null hypothesis. Since the inquiry concerns a hypothetical deterrent effect, the null hypothesis should be constructed in a form that permits testing of the relevant set of behavioral relations implied by a general theory of deterrence. That includes the deterrent effects of law enforcement activities in general. Moreover, even if a negative effect of capital punishment on the rate of murder is established through systematic empirical research, there still remains the question of the existence of a *pure* deterrent effect distinct from a potential preventive or incapacitating effect associated with this form of punishment. An effect of the second type might be expected since execution eliminates categorically the possibility of recidivism.

Contrary to previous observations, this investigation, although by no means definitive, does indicate the existence of a pure deterrent effect of capital punishment. In fact, the empirical analysis suggests that on the average the tradeoff between the execution of an offender and the lives of potential victims it might have saved was of the order of magnitude of 1 for 8 for the period 1933–67 in the United States. Two related arguments are offered in this context of which only the second will be elaborated in this paper. First, it may be argued that the statistical methods used by Sellin and others to infer the nonexistence of the deterrent effect of capital punishment do not provide an acceptable test of such an effect and consequently do not warrant such inferences. Second, it is argued that the application of the economic approach to criminality and the identification of relevant determinants of murder and their empirical counterparts permit a more systematic test of the existence of a differential deterrent effect of capital punishment. The theoretical approach, emphasizing the interaction between offense and defense—the supply of and the (negative) social demand for murder—is developed in Section I. Section II is devoted to the empirical implementation of the model. Some implications of the empirical evidence are then presented and discussed in Section III.

I. An Economic Approach to Murder and Defense Against Murder

A. *Factors Influencing Acts of Murder and Other Crimes Against Persons*

The basic propositions underlying the approach to murder and other crimes against the person are 1) that these crimes are committed largely as a result of hate, jealousy, and other interpersonal conflicts involving pecuniary and nonpecuniary motives or as a by-product of crimes against property; and 2) that the propensity to perpetrate such crimes is influenced by the prospective gains and losses associated with their commission. The abhorrent, cruel, and occasionally pathological nature of murder notwithstanding, available evidence is at least not inconsistent with these basic propositions. Victimization data reveal that most murders, as well as other crimes against the person, occur within the family or among relatives, friends, and other persons previously known to one another, and are not committed as a rule by strangers on the street (see President's Commission on Law Enforcement and Administration of Justice (*PCL*), pp. 14, 15, 81, and 82). Indeed, hate and other interdependencies in utility across persons as well as malevolent and benevolent exchanges would seem more likely to develop among groups that exercise relatively close and frequent social contact than among groups that exercise little or no contact. There is no reason a priori to expect that persons who hate or love others are less responsive to changes in costs and gains associated with activities they may wish to pursue than persons indifferent toward the well-being of others.

More formally, assume that person o's utility from a consumption prospect C_o, depends upon his own consumption c_o, and consumption activities involving other persons c_j, $j=1, \ldots n$, or

$$(1) \qquad U_o(C_o) = U_o(c_o, c_j)$$

where the sign of $\partial U_o/\partial c_j$ indicates the direction in which o's utility is affected by consumption activities pursued by others. The key feature of this consumption model involving interdependent preferences[1] is that it provides a framework for analyzing positive or negative transfers of resources by one person (here identified with person o) that modify the levels of consumption enjoyed by others while simultaneously determining his own consumption level. Such modifications are constrained generally by the pertinent transfer production functions, by the endowments of resources possessed by person o and other relevant persons, and by potential awards and penalties that are conditional upon o's benevolent or malevolent actions with varying degrees of uncertainty.[2]

This framework can be applied to analysis of the incentive to commit murder and other crimes against the person by explicitly incorporating into the model the uncertainties associated with the prospective punishments for crime. Specifically, murder can be considered a deliberate action intended by an offender o to inflict severe harm on a victim v by setting c_v equal to, say, zero. The offender undergoes some direct costs of planning and executing the crime, and bears the risk of incurring detrimental losses in states of the world involving apprehension, conviction, and punishment.[3] Assuming the offender

---

[1] For a more complete discussion of this model, see Harold Hochman and James Rodgers, and Gary Becker (1974).

[2] It might be argued that although the wish to harm other persons cannot be rejected on economic grounds, nonetheless the execution of such desires (as opposed to benevolent actions) must be considered irrational in the sense of violation of Pareto optimality conditions. If there were no bargaining, transfer, or enforcement costs associated with mutually acceptable and enforceable contracts between a potential offender o and his potential victim v, and if v's wealth constraint were not binding, then it would always be optimal for v to offer compensation to o for not committing a crime against him and for o to seek such compensation or extortion. The reason is that a reduction in v's consumption level is thus achieved by o without incurring the direct costs of committing a crime and the prospective cost of legal sanctions. Indeed, there exists some range of compensations that would increase both o's and v's utilities relative to their expected utilities if crime is committed by o against v. Many crimes against persons, and some cases of property crimes as well, may occasionally be avoided by such arrangements; successful extortions involving kidnapping or hijacking constitute obvious examples. Yet in many situations compensations may be too costly to pursue or to enforce, just as fully effective private or public protection against murder may be too costly to provide. The incidence of murder must then be expected on purely economic grounds.

[3] The case in which crime is committed in pursuit of material gains has been analyzed explicitly by the

TABLE 1

Event		State s	Probabilities π_s	Consumption Prospect C_s
Apprehension	conviction of murder	execution	$(Pa)(Pc\|a)(Pe\|c)$	C_d: $(c_o=0;\ c_v=0)$
		imprisonment for murder	$(Pa)(Pc\|a)(1-Pe\|c)$	C_c: $(c_o=c;\ c_v=0)$
	conviction of a lesser offense or acquittal	other punishment	$Pa(1-Pc\|a)$	C_b: $(c_o=b;\ c_v=0)$
No Apprehension		no punishment	$1-Pa$	C_a: $(c_o=a;\ c_v=0)$

behaves as if to maximize expected utility, a necessary and sufficient condition for murder to occur is that o's expected utility from crime exceeds his expected utility from an alternative (second best) action:

$$(2)\quad U^*_{om}(C^m_o \mid c_v = 0) \equiv \sum_{s=a}^{S} \pi_s U_o(C_{os}) > U^*_{ol}(C^l_o \mid c_v = c^l_v),$$

where $s=a, \ldots, S$ denote a set of mutually exclusive and jointly exhaustive states of the world including all the possible outcomes of murder; c_{os} denote the offender's consumption levels, net of potential punishments and other losses, that are contingent upon these states; π_s denote his subjective evaluation of the probabilities of these states; and C^m_o and C^l_o denote, respectively, his consumption prospect in the event he commits murder or takes an alternative action.

To illustrate the behavioral implications of the model via a simple yet sufficiently general example, assume the existence of just four states of the world associated with the prospect of murder as summarized in Table 1. In Table 1, Pa denotes the probability of the event of apprehension and $1-Pa$ denotes its complement—the probability of escaping apprehension; $Pc|a$ denotes the conditional probability of conviction of murder given apprehension, and $1-Pc|a$ denotes its complement—the probability of conviction of a lesser offense (including acquittal); finally, $Pe|c$ and $1-Pe|c$ denote, respectively, the conditional probabilities of execution and of other punishments given conviction of murder. The (subjective) probabilities of the set of states introduced in Table 1 are equal by definition to the relevant products of conditional probabilities of sequential events that lead to this more final set of states. The last column in Table 1 lists the consumption levels that are contingent upon the occurrence of these states. Economic intuition suggests that the relevant consumption levels can be ranked according to the severity of punishment imposed on the offender; that is, $C_a > C_b > C_c > C_d$.

In the preceding discussion the incidence of murder has been viewed to be motivated by hate. As hinted earlier in the discussion, however, murder could also be a by-product, or more generally, a complement of other crimes against persons and property. Since the set of states of the world underlying the outcomes of these other crimes also includes punishment for murder, the decision to commit these would also be influenced by factors determining the probability distribution of outcomes considered in Table 1. In turn, the incidence of murder would be influ-

---

author (1973a). Note that the victim's level of consumption need not directly enter the offender's utility function in this case.

enced by factors directly responsible for related crimes. In general, behavioral implications concerning the effect of various opportunities on the incidence of murder ought to be analyzed within a framework that includes related crimes as well. For methodological simplicity and because data exigencies rule out a comprehensive empirical implementation of such a framework, the following discussion emphasizes the effect of factors directly related to murder and the direct effect on murder of general economic factors like income and unemployment. In practice, however, the effect of these latter factors on murder may be due largely to their systematic effects on particular crimes against property.

1. *The Effects of Probability and Severity of Punishment*

An immediate implication of the model that is independent of the specific motives and circumstances leading to an act of murder is that an increase in the probability or severity of various punishments for murder decreases, relative to the expected utility from an alternative independent activity, the expected utility from murder or from activities that may result in murder. These implications have been discussed at length elsewhere (see the author (1970, 1973a)) but the somewhat more detailed formulation of the model adopted in this paper makes it possible to derive more specific predictions concerning the relative magnitudes of the deterrent effects of apprehension, conviction, and execution that expose the theory to a sharper empirical test. Specifically, given the ranking of the consumption levels in states of the world involving execution, imprisonment, other punishment, and no punishment for murder, as assumed in the preceding illustration, and given the level of the probabilities of apprehension and the conditional probabilities of conviction and execution, it can be shown that the partial elasticities of the expected utility from crime with respect to these probabilities can be ranked in a descending order as follows:

$$\epsilon_{Pa} > \epsilon_{Pc|a} > \epsilon_{Pe|c} \tag{3}$$

where $\epsilon_P = -\partial \ ln \ U^*/\partial \ ln \ P$ for $P = Pa$, $Pc|a$, $Pe|c$.[4] The interesting implication of condition (3) is that the more general the event leading to the undesirable consequences of crime, the greater the deterrent effect associated with its probability: a 1 percent increase in the (subjective) probability of apprehension Pa, given the values of the conditional probabilities $Pc|a$ and $Pe|c$, reduces the expected utility from murder more than a 1 percent increase in the conditional (subjective) probability of conviction of murder $Pc|a$ (as long as $Pc|a<1$), essentially because an increase in Pa increases the overall, i.e., unconditional, probabilities of three undesirable states of the world: execution, other punishment for murder, and punishment for a lesser offense, whereas an increase in $Pc|a$ raises the unconditional probability of the former two states only. A fortiori, a 1 percent increase in $Pc|a$ is expected to have a greater deterrent effect

---

[4] Differentiating equation (2) with respect to Pa, $Pc|a$, and $Pe|c$, using the contingent outcomes of murder as illustrated in Table 1, it can easily be demonstrated that:

$$\begin{aligned}
\epsilon_{Pa} &= -\frac{\partial U^*_{om}}{\partial Pa}\frac{Pa}{U^*_o} = \frac{1}{U^*_o}\{Pa(1-Pc|a) \\
&\quad \cdot[U(C_a)-U(C_b)] + PaPc|a(1-Pe|c) \\
&\quad \cdot[U(C_a)-U(C_c)] + PaPc|aPe|c \\
&\quad \cdot[U(C_a)-U(C_d)]\} > 0 \\
\epsilon_{Pc|a} &= -\frac{\partial U^*_{om}}{\partial Pc|a}\frac{Pc|a}{U^*_o} = \frac{1}{U^*_o}\{PaPc|a(1-Pe|c) \\
&\quad \cdot[U(C_b)-U(C_c)] + PaPc|aPe|c \\
&\quad \cdot[U(C_b)-U(C_d)]\} > 0 \\
\epsilon_{Pe|c} &= -\frac{\partial U^*_{om}}{\partial Pe|c}\frac{Pe|c}{U^*_o} = \frac{1}{U^*_o}\{PaPc|aPe|c \\
&\quad \cdot[U(C_c)-U(C_d)] > 0
\end{aligned}$$

Clearly, $\epsilon_{Pa} > \epsilon_{Pc|a} > \epsilon_{Pe|c} > 0$.

than a 1 percent increase in $Pe|c$ as long as $Pe|c$ is less than unity. If there exists a positive monotonic relation between an average person's subjective evaluations of Pa, $Pc|a$, and $Pe|c$ and the objective values of these variables, and between an average person's expected utility from crime and the actual crime rate in the population, equation (3) would then amount to a testable theorem regarding the partial elasticities of the murder rate in a given period with respect to objective measures of Pa, $Pc|a$, and $Pe|c$. On the basis of this analysis, it can be predicted that while the execution of guilty murderers deters acts of murder, *ceteris paribus*, the apprehension and conviction of guilty murderers is likely to have an even larger deterrent effect.

Analogous to the effects of the probabilities of various punishments for murder, an increase in the severity of these punishments, their probabilities held constant, is generally expected to decrease the expected utility from murder and so to discourage its commission. Due to lack of space, other implications concerning the effect of severity as well as probability of punishment on the elasticities ϵ_{Pa}, $\epsilon_{Pc|a}$, and $\epsilon_{Pe|c}$ are omitted here. For a more complete analysis, see the author (1973b).

2. *Effects of Employment Opportunities, Income, and Demographic Variables*

The model developed in this section suggests that the incentive to commit murder or other crimes that may result in murder in general would depend on permanent income (or wealth), the relevant opportunities to extract related material gains as well as on direct opportunities for malevolent actions, including the direct costs involved in effecting the production of malevolent transfers. The means for a direct implementation of the effect of these latter opportunities are not readily available (see, however, the discussion in fn. 14). In contrast, variations in legitimate and illegitimate earning and income opportunities may be approximated by movements in the rate of unemployment and of labor force participation, U and L, respectively, and in the level and distribution of permanent income Y_p in the population.

The relevance of the latter set of variables has been discussed in detail elsewhere (see the author (1973a)), particularly in connection with crimes against property, some of which involve murder. However, the level and distribution of income within a community may also exert a direct influence on the incentive to commit murder because of their impact on the individual demand for malevolent actions. In addition, although the decision to commit murder is presumably derived from considerations related to lifetime utility maximization, the timing of murder may be affected by variations in the opportunity cost of time throughout the life cycle, because the typical punishment for murder involves a finite imprisonment term. Thus, to the extent that earning opportunities are imperfectly controlled in an empirical investigation, it may be important to investigate the independent effects of variations in demographic variables, such as the age and racial composition of the population, A and NW, respectively. Controlling for variations in age composition may also be important because of the differential treatment of young offenders under the law.

B. *Defense Against Murder*

1. *Factors Determining Optimal Law Enforcement Activity*

Following the approach used by Becker (1968), I shall attempt to derive implications concerning law enforcement activity against murder on the assumption that law enforcement agencies behave as if they seek to maximize a social welfare function by minimization of the per capita loss from murder. Losses accrue from three main elements: harm to victims net of gains to offenders; the direct costs of law enforcement by police and courts; and the net social costs associated with penalties. The behavior of enforcement agencies is assumed to be in accordance with the general implications of the deterrent theory of law enforcement.

The main elements of the social loss function can be summarized by:

$$L = D(q) + C(q, Pc) + \gamma_1 Pc\, Pe \,|\, c\, qd + \gamma_2 Pc(1 - Pe \,|\, c)qm \tag{4}$$

The term $D(q)$ represents the net social damage resulting from the death of victims and other related losses, where $q \equiv Q/N$ denotes the rate of murder in the population. The term $C(q, Pc)$ represents the total cost of apprehending, indicting, prosecuting, and convicting offenders. The aggregate output of these law enforcement activities can be summarized by the fraction of all murders that are "cleared" by the conviction of their alleged perpetrators (assuming a fixed proportional relation between the number of murders and their perpetrators). This fraction θ may be viewed as an objective indicator of the probability that a perpetrator of murder will be convicted of his crime, $Pc = Pa(Pc \,|\, a)$ with one qualification: since the overall probability of error of justice, π—that of apprehending and convicting an innocent person—is greater than nil, the true probability of conviction $0 < Pc < 1$ will be systematically lower than θ. However, to abstract the analysis from a separate determination of the optimal value of ϵ, it is henceforth assumed that Pc and θ are proportionally related, so that C can be defined as a direct function of Pc.[5] The rate of murder q is introduced as a separate determinant of C because of the argument and evidence that the costs of producing a given value of θ are higher for higher levels of q. The larger is q, the larger the number of suspects that must be apprehended, charged, and convicted in order to achieve a given value of θ. Both D and C are assumed to be monotonically increasing, continuously differentiable, and concave functions in each of their respective arguments.

The third and fourth terms in equation (4) represent the per capita social costs of punishing guilty and innocent convicts through execution and imprisonment (or other penalties), respectively. The variables d and m denote the private costs to victims and their families from execution and imprisonment, and the multipliers γ_1 and γ_2 indicate the presence of additional costs or gains to the rest of society from administering and otherwise bearing the respective penalties of execution and imprisonment that are imposed on guilty and innocent convicts.[6] For methodological convenience, the costs of execution and imprisonment can be combined, and equation (4) can be rewritten as:

---

[5] Pc and θ would be proportionally related if the number of arrests of innocent and guilty persons were proportionally related and if the probability of legal error remained constant as more resources were spent on enforcement activity through arrests and prosecutions. Alternatively, it might be argued that Pc and θ are highly (positively) correlated because of the well-known proposition that at any given level of evidence presented in court in reference to the defendant's guilt or innocence, the probability of legal or type I error, α (that of convicting the innocent), is negatively related to the probability of type II error, β (that of acquitting the guilty). Hence α might be negatively correlated with $Pc \,|\, ch = 1 - \beta$ where $P_c \,|\, ch$ denotes the conditional probability that a guilty offender will be convicted once he is charged. However, the assumption that Pc and θ, or $Pc \,|\, ch$ and α, are mutually dependent is made mainly for methodological convenience without affecting the basic implications of the following analysis. More generally, the direct costs of law enforcement activity C may be specified as a function including Pc and the unconditional probability of legal error ϵ as independent arguments so that optimal values of these probabilities may be determined separately via appropriate expenditures.

[6] More specifically, $\gamma_1 = b_1 + \lambda\beta_1$ and $\gamma_2 = b_2 + \lambda\beta_2$, where λ is a coefficient relating Pc to the fraction of murders cleared by convicting innocent persons π and b and β indicate the respective net social costs from

$$(5) \qquad L = D(q) + C(q, Pc) + \gamma_1 Pcfq$$

where $f = (Pe|c)d + \gamma_2(1 - Pe|c)m/\gamma_1$ is a measure of the average social cost of punishment for murder.

Equation (5) identifies the unconditional probability of conviction Pc, and the expected social cost of punishment f, as the main control variables underlying law enforcement activity. Given the harshness of the method of execution, the length of imprisonment terms, and other factors determining d and m (changes in these factors occur slowly in practice) the magnitude of f is largely a function of the conditional probability of execution $Pe|c$. The values of $0 < Pc < 1$ and $0 < Pe|c < 1$ that locally minimize equation (5) must then satisfy the following pair of equilibrium conditions:

$$(6) \quad \left[D_q + C_q + C_p \frac{1}{q_p} + \gamma_1 Pcf(1 - Ep) \right] q_p = 0$$

$$(7) \qquad [D_q + C_q + \gamma_1 Pcf(1 - Ef)] q_f f_e = 0$$

where

$$Ep \equiv - \frac{\partial Pc}{\partial q} \frac{q}{Pc} \equiv \frac{1}{\varepsilon_{Pc}}$$

$$Ef \equiv - \frac{\partial f}{\partial q} \frac{q}{f} \equiv \frac{1}{\varepsilon_f}$$

$$f_e \equiv \frac{\partial f}{\partial Pe|c} = \left(d - \frac{\gamma_2}{\gamma_1} m \right)$$

and the subscripts p, f, and e associated with the variables C and q denote the partial derivatives of the latter with respect to Pc, f, and $Pe|c$, respectively. The product $\gamma_1 f_e$ indicates the difference between the social costs of execution and imprisonment.

In equation (7) the term $-(D_q + C_q)q_f f_e$ represents part of the marginal revenue from execution: the value of the lives of potential victims saved, and the reduced costs of apprehending and convicting offenders due to the differential deterrent effect of execution on the frequency of murder. The term $\gamma_1 Pcf(1 - Ef)q_f f_e$ represents the net marginal social cost of execution: the value to society of the life of a person executed at a given probability of legal error, plus all the various costs of effecting his execution (including mandatory appeals) net of imprisonment costs thereby "saved." Because in equilibrium, the two must be equated, the optimal value of $Pe|c$ need not be unity—capital punishment may not always be imposed even when it is legal—and would depend on the relative magnitude of the relevant costs and gains. A similar interpretation applies to equation (6).

Inspection of the equilibrium conditions given by equations (6) and (7) reveals a number of interesting implications. First, it may be noted that if an increase in $Pe|c$ unambiguously raises the social cost of punishment for murder, that is, if $\gamma_1 f_e = \gamma_1 d - \gamma_2 m > 0$, then in equilibrium, the deterrent effect associated with capital punishment must be less than unity, or $\varepsilon_{P|e} < \varepsilon_f < 1$.[7] Put differently, executions must only decrease the rate of murders in the population but not the rate of persons executed, for otherwise the marginal cost of execution would be negative and a corner solu-

---

punishing guilty and innocent convicts through execution, denoted by the subscript 1, or imprisonment, denoted by the subscript 2. The conditional probability of execution given conviction is implicitly assumed to be equal for all convicts.

[7] By definition,

$$\varepsilon_{Pe|c} \equiv - (\partial q / \partial Pe|c)(Pe|c/q)$$
$$\equiv \varepsilon_f (\partial f / \partial Pe|c)(Pe|c/f) \equiv \varepsilon_f \varepsilon_{f_e}$$

Clearly,

$$\varepsilon_{f_e} = Pe|c[d - (\gamma_2/\gamma_1)m] / \{(\gamma_2/\gamma_1)m + Pe|c[d - (\gamma_2/\gamma_1)m]\}$$

is lower than unity if $[d - (\gamma_2/\gamma_1)m] > 0$. Under this condition, and the assumption that $\gamma_1 > 0$, $\varepsilon_{Pe|c} < \varepsilon_f < 1$.

tion would be achieved at $Pe|c=1$. However, equation (7) does not have the same implications regarding the value of ε_{Pc}. More specifically, equation (6) shows that the marginal costs of conviction include the marginal costs of apprehending and convicting offenders, in addition to the marginal costs of punishing those convicted. Therefore, the overall marginal revenue from convictions must also be higher than that from executions. Indeed, by combining equations (6) and (7), it can readily be shown that in equilibrium, $\varepsilon_{Pc} > \varepsilon_f > \varepsilon_{Pe|c}$;[8] that is, the deterrent effect associated with Pc must exceed the differential deterrent effect associated with $Pe|c$. This proposition is essentially the same as that derived regarding the response of offenders to changes in Pc and $Pe|c$ (see equation (3)). The compatibility of the implications of optimal offense and defense under the assumption that both offenders and law enforcement agencies regard execution to be more costly than imprisonment insures the stability of equilibrium with respect to both activities. It also provides the basis for a sharp empirical test of the theory.

2. *The Interdependencies Among the Murder Rate and the Probabilities of Conviction and Execution*[9]

Any exogenous factor causing a decrease in the severity of punishment for murder via a decrease in $Pe|c$ can be shown to increase the value of Pc because it tends to decrease the marginal costs of conviction and increase its marginal revenue. More specifically, given the values of d and m, an increase in social aversion toward capital punishment or in the costs of the related due process, measured by γ_1, can be shown to produce a decline in the optimal value of $Pe|c$ and a simultaneous increase in the optimal value of Pc. This analysis is consistent with an argument often made regarding the greater reluctance of courts or juries to convict defendants charged with murder when the risk of their subsequent execution is perceived to be undesirably high. Conviction and execution thus can be considered substitutes in response to changes in the shadow price of each. Indeed, the empirical investigation reveals that at least over the period between 1933 and 1969, in which the estimated annual fraction of convicts executed for murder in the United States, denoted by PXQ_1, fell from roughly 8 percent to nil, the national clearance ratios of reported murders, denoted by P^0a, and the fraction of persons charged with murder who were convicted of murder, denoted by $P^0c|a$, on the whole, moved in an opposite direction. The zero order correlation coefficient between PXQ_1, and P^0a is found to be -0.028, while that between PXQ_1 and $P^0c|a$ is found to be -0.19. (In principle, the product $P^0aP^0c|a$ approximates the value of Pc.) The general implication of this analysis is that the simple correlation between estimates of the murder rate and the conditional probabilities of execution cannot be accepted as an indicator of the true differential deterrent effect of capital punishment, because the simple correlation is likely to confound the offsetting effects of opposite changes in Pc and possibly also in the probability and severity of alternative punishments for murder.

Just as convictions and executions are expected to be substitutes with respect to changes in the shadow cost of each activity, they can be expected to be complementary with respect to changes in the severity of damages from crime, essentially because such changes increase the marginal revenues from both activities. Since an exogenous increase in the rate of murder is expected to increase the marginal social damage D_q, and, indirectly, the marginal costs of apprehension and conviction C_q, it is expected to induce an increase in the optimal values of both Pc and $Pe|c$. This analysis demonstrates the simultaneous relations between offense and defense and suggests that the deterrent effects of conviction and execution must be identified empirically through appropriate simultaneous equation estimation techniques.

---

[8] By like reasoning and some simplifying assumptions, it can also be shown that in equibrium, $\varepsilon_{Pa} > \varepsilon_{Pe|a} > \varepsilon_{Pc|a}$.

[9] Proofs to the theorems discussed in this section can be developed through an appropriate differentiation of equations (6) and (7) with respect to the relevant variables.

II. New Evidence on the Deterrent Effect of Capital Punishment

A. *The Econometric Model*

In the empirical investigation an attempt is made to test the main behavioral implications of the theoretical model. The econometric model of crime and law enforcement activity devised by the author (1973a) is applied to aggregate crime statistics relating to the United States for the period 1933–69. The model treats estimates of the murder rate and the conditional probabilities of apprehension, conviction, and execution as jointly determined by a system of simultaneous equations. Since data limitations rule out an efficient estimation of structural equations relating to law enforcement activities or private defense against murder, the following discussion focusses on a supply-of-murders function actually estimated in this study.

1. *The Murder Supply Function*

It is assumed that the structural equations explaining the endogenous variables of the model are of a Cobb-Douglas variety in the arithmetic means of all the relevant variables. The murder supply function is specified as follows.

$$(8)\quad \left(\frac{Q}{N}\right) = CPa^{\alpha_1}Pc\,|\,a^{\alpha_2}Pe\,|\,c^{\alpha_3}U^{\beta_1}L^{\beta_2}Y_p^{\beta_3}A^{\beta_4}\exp(v_1)$$

where C is a constant term and v_1 is a disturbance term assumed to be subject to a first-order serial correlation. The regression equation thus can be written as:

$$(9)\qquad y_1 = Y_1A_1' + X_1B_1' + v_1$$

where

$$(10)\qquad v_1 = \rho v_{1_{-1}} + e_1$$

The variables y_1, Y_1, and X_1 denote, respectively, the natural logarithms of the dependent variable, other endogenous variables, and all the exogenous variables entering equation (8); ρ denotes the coefficient of serial correlation, and the subscript -1 denotes one-period lagged values of a variable. The coefficient vectors A_1' and B_1' have been estimated jointly with ρ and the standard error of e_1, σ_e, via a nonlinear three-round estimation procedure proposed by Ray Fair.

2. *Variables Used*

The dependent variable of interest (Q/N) is the true rate of capital murders in the population in a given year. The statistic actually used, $(Q/N)^0$, is the number of murders and nonnegligent manslaughters reported by the police per 1,000 civilian population as computed from data reported by the FBI *Uniform Crime Report* (*UCR*)[10] and the Bureau of the Census. This statistic can serve as an efficient estimator of the true Q/N if the two were related by:

$$(11)\qquad \left(\frac{Q}{N}\right) = k\left(\frac{Q}{N}\right)^0 \exp\mu$$

where k indicates the ratio of the true number of capital murders committed in a given year relative to all murders reported to the police, and μ denotes random errors of reporting or identifying murders. It should be noted, however, that the fraction of capital murders among all murders may have been subject to a systematic trend over time. Indeed, the theory developed in Section IA suggests that the decrease in the tendency to apply the death penalty in the United States over time may have led to an increase in the fraction of capital murders among all murders. More important, the number of reported murders may have decreased systematically over time because of the decrease in the fraction of all attempted murders resulting in the death of the victims due to the continuous improvement in medical technology. To account for such possible trends, the term k in equation (11) can be defined as $k=\delta \exp(\lambda T)$, where δ and λ are constant terms and T denotes chronological time. Upon substitution of $(Q/N)^0$ for (Q/N) in equation (8), the inverse values of δ and μ would be subsumed under the constant term C and the stochastic variable v, respectively, and $\exp(-\lambda T)$ would emerge as an additional explanatory variable. Thus, the natural value of T is introduced in equation (9) as an independent exogenous variable.[11]

---

[10] I am indebted to the Uniform Crime Reporting Section of the FBI for making available their revised annual estimates of the total number of murders and other index crimes in the United States during the period 1933–65.

The matrix of endogenous variables associated with Y_1 in equation (9) includes the conditional probabilities that guilty offenders be apprehended, convicted, and executed for murder. These probabilities have been approximated by computing objective measures of the relevant fractions of offenders who are apprehended, convicted, and executed. The following paragraphs contain a brief discussion of these measures.

Pa is measured by the national "clearance rates" as reported by the FBI *UCR*, which are estimates of the percentage of all murders cleared by the arrest of a suspect. It is denoted by P^0a. The conditional probability $Pc|a$ is identically equal to $Pch|a \cdot Pc|ch$—the product of the conditional probabilities that a person who committed murder be charged once arrested, and that he be convicted once charged. Statistical exigencies preclude the estimation of a complete series of $Pch|a$, but $Pc|ch$ is estimated by the fraction of all persons charged with murder who were convicted of murder in a given year as reported by the FBI *UCR*. This fraction, denoted by $P^0c|a$, may serve as an efficient estimator of the overall true probability $Pc|a$, provided that $Pch|a$ were either constant over time, or proportionally related to the probability of arrest Pa.

The actual measures of $Pe|c$ consist of alternative estimates of the expected fractions of persons convicted of murder in a given year who were subsequently executed, $P^0e|c$. Because no complete statistics on the disposition of murder convicts by type of punishment are available, however, $P^0e|c$ has been estimated indirectly by matching annual time-series data on convictions and executions. Over most of the period considered in this investigation (up to 1962), executions appear to lag convictions by 12 to 16 months on the average. An objective measure of $P^0e|c$ in year t, therefore, may be the ratio of the number of persons executed in year t+1 to the number convicted in year t or $PXQ_1=E_{t+1}/C_t$.[12] It must be pointed out, however, that the number of persons executed in year t+1, and hence PXQ_1, is, of course, unknown in year t and must be forecast by potential offenders. Even if expectations with respect to PXQ_1 were unbiased on the average, the actual magnitude of PXQ_1 is likely to deviate randomly from its expected value in year t. The effect of such random noise would be to bias the regression coefficient associated with PXQ_1 toward zero. I have therefore constructed four alternative forecasts of the desired variable: $PXQ_{1_{-1}} = E_t/C_{t-1}$; $PXQ_2 = E_t/C_t$; $TXQ_1 =$ the systematic part of PXQ_1 computed via a linear distributed lag regression of PXQ_1 on three of its consecutively lagged values; and $PDL_1 =$ the systematic part of PXQ_1 computed via a second degree polynomial distributed lag function relating PXQ_1 and four of its consecutively lagged values. The advantage of using these alternative estimates of the expected $P^0e|c$ is that all being based on past data, they may be treated largely as predetermined rather than as endogenous variables. Alternatively, PXQ_1 is treated as an endogenous variable along with P^0a and $P^0c|a$, and its systematic part is computed via the reduced form regression equation (see Table 3).

---

[11] Another important reason for introducing chronological time as an exogenous variable in equation (8) is to account for a possible time trend in missing variables, in particular, the average length of imprisonment for both capital and noncapital murders for which no complete time-series is available. Scattered evidence shows rising trends in the median value of prison terms served by all murder convicts over a large part of the period considered in this investigation, but this increase may have been largely technical. With executions being imposed less frequently over time, the frequency of life imprisonment sentences for murder convicts may have risen accordingly, thus increasing the mean or median time spent in prisons by these convicts.

[12] Execution figures are based on *National Prisons Statistics Bulletin* (*NPS*) statistics. Conviction figures are derived by $C_t=Q_t^0P^0a_tP^0c|a_t$. Statistics on the time elapsed between sentencing and execution can be found in *NPS* numbers 20 and 45.

Two difficulties associated with the use of the proposed estimates of $P^0e|c$ as measures of the true conditional probability of execution warrant special attention. First, it may be argued that the fraction of convicts executed for murder may represent only the fraction of those convicted of *capital* murders among all murder convicts. Variations in PXQ_1 or in other related estimates might then be entirely unrelated to the probability that a convict liable to be punished by the death penalty will be actually executed, and the expected elasticity of the murder rate with respect to these estimates might be nil. However, the significant downward trend in PXQ_1 between 1933 and 1967 suggests, especially during the 1960's, that it may serve as a useful indicator of the relative variations in the true $Pe|c$. Second, it should be noted that the relative variation in the reported national murder rate relates to the United States as a whole, whereas the measures of $P^0e|c$ relate to only a subset of states which retained and actually enforced capital punishment throughout the period considered. Thus, the empirical estimates of the elasticities of the national murder rate with respect to $P^0e|c$ may, on this ground, be expected to understate the true elasticities of the murder rate in retentionist states only.

The matrix of exogenous variables associated with X_1 in equation (9) includes annual census estimates of the labor force participation rate of the civilian population 16 years and over (calculated by excluding the armed forces from the total noninstitutional population) L; the unemployment rate of the civilian labor force U; Milton Friedman's estimate of real per capita permanent income (extended through 1969)[13] Y_p; the percentage of residential population in the age group 14–25, A; and chronological time T. Other exogenous variables assumed to be associated with the complete simultaneous equation model of murder and law enforcement X_2 are one-year lagged estimates of real expenditure on police per capita $XPOL_{-1}$ and annual estimates of real expenditure by local, state, and federal governments per capita $XGOV$. Real expenditures are computed by deflating *Survey of Current Business* estimates of current expenditures by the implicit price deflator for all governments. In addition, X_2 includes the size of the total residential population in the United States N, and the percent of nonwhites in residential population NW. The reason for including NW in the list of variables subsumed under X_2 is discussed below in Section IIB. A list of all the variables used in the regression analysis is given in Table 2.

---

[13] I am indebted to Edi Karni for making available to me his updated calculations of the permanent income variable.

B. *The Empirical Findings*

An interesting finding which poses a challenge to the validity of the analysis in Section I is that over the period 1933–69, the simple correlation between the reported murder rate and estimates of the objective risk of execution given conviction of murder is positive in sign. For example, the simple correlation coefficients between $(Q/N)^0$ and PXQ_1, $PXQ_{1_{-1}}$, and PXQ_2 are found to be 0.140, 0.096, and 0.083, respectively. However, the results change substantively and are found to be in accordance with the theoretical predictions and statistically meaningful when the full econometric framework developed in the preceding section is implemented against the relevant data. Despite the numerous limitations inherent in the empirical counterparts of the desired theoretical constructs, the regression results reported in Tables 3 and 4 uniformly exhibit a significant negative elasticity of the murder rate with respect to each alternative measure of the probability of execution. More importantly, the regression results also corroborate the specific theoretical predictions regarding the effects of apprehension, conviction, unemployment, and labor force participation.

Table 3 shows that the estimated elasticity of the murder rate with respect to the conditional probability of execution is

TABLE 2—VARIABLES USED IN THE REGRESSION ANALYSIS, ANNUAL OBSERVATIONS 1933–69

	Variable	Mean (Natural Logarithms)	Standard Deviation (Natural Logarithms)	Arithmetic Mean
y_1	$(Q/N)^0$ = Crime rate: offenses known per 1,000 civilian population.	−2.857	0.156	0.058
Y_1	P^0a = Probability of arrest: percent of offenses cleared.	4.997	0.038	89.835
	$P^0c\|a$ = Conditional probability of conviction: percent of those charged who were convicted of murder.[a]	3.741	0.175	42.733
	$P^0e\|c$ = Conditional probability of execution; PXQ_1 = the number of executions for murder in the year t+1 as a percent of the total number of convictions in year t.[b]	0.176	1.749	2.590
X_1	L = Labor force participation: fraction of the civilian population in the labor force.	−0.546	0.030	0.579
	U = Unemployment rate: percent of the civilian labor force unemployed.	1.743	0.728	7.532
	A = Fraction of residential population in the age group 14–24.	−1.740	0.118	0.177
	Y_p = Friedman's estimate of (real) permanent income per capita in dollars.	6.868	0.338	1012.35
	T = Chronological time (years): 31–37.	2.685	0.867	19.00
X_2	NW = Fraction of nonwhites in residential population.	−2.212	0.063	0.110
	N = Civilian population in 1,000s.	11.944	0.161	155,853
	$XGOV$ = Per capita (real) expenditures (excluding national defense) of all governments in million dollars.	−7.661	0.501	.000532
	$XPOL_{-1}$ = Per capita (real) expenditures on police in dollars lagged one one year.[a]	2.114	0.306	8.638

[a] The figures for $P^0c|a$ (1933–35) and $XPOL$ (all the odd years 1933–51) were interpolated via an auxiliary regression analysis.

[b] The actual number of executions 1968, 1969, and 1970 was zero. However, the numbers were assumed equal to 1 in each of these years in constructing the value of PXQ_1 in 1967–69.

TABLE 3—MODIFIED FIRST DIFFERENCES OF MURDER RATES (IN NATURAL LOGARITHMS) REGRESSED AGAINST CORRESPONDING MODIFIED FIRST DIFFERENCES OF SELECTED VARIABLES SET I (1933–69)

($\hat{\beta}/S_{\hat{\beta}}$ in parentheses)

Effective Period *D.W.* Statistic	$\hat{\rho}$(*CORC*) $\hat{\sigma}_e$	*C* (Constant)	$\Delta^*\hat{P}^o a$	$\Delta^*\hat{P}^o c\|a$	Alternative $\Delta^* P^o e\|c$ Δ^*PXQ_1	 Δ^*PXQ_2	 Δ^*PXQ_{1-1}	Δ^*L	Δ^*A	Δ^*Y_p	Δ^*U	Δ^*T
1. 1935–69	0.257	−3.176	−1.553	−0.455	−0.039			−1.336	0.630	1.481	0.067	−0.047
1.84	0.052	(−0.78)	(−1.99)	(−3.58)	(−1.59)			(−1.36)	(2.10)	(4.23)	(2.00)	(−4.60)
2. 1935–69	0.135	−4.190	−1.182	−0.386		−0.068		−1.277	0.481	1.318	0.062	−0.047
1.82	0.042	(−1.25)	(−1.83)	(−3.85)		(−3.69)		(−1.59)	(2.19)	(4.86)	(2.38)	(−6.61)
3. 1935–69	0.077	−4.419	−1.203	−0.374			−0.065	−1.405	0.512	1.355	0.068	−0.047
1.81	0.045	(−1.25)	(−1.78)	(−3.59)			(−3.29)	(−1.63)	(2.26)	(4.88)	(2.55)	(−6.54)
					Δ^*TXQ_1	Δ^*PDL_4	$\Delta^*P\hat{X}Q_1$					
4. 1937–69	0.291	−2.447	−1.461	−0.487	−0.049			−1.393	0.524	1.295	0.063	−0.044
2.00	0.046	(−0.61)	(−2.03)	(−3.38)	(−2.26)			(−1.58)	(1.94)	(3.90)	(2.09)	(−4.93)
5. 1939–69	−0.207	6.868	−2.225	−0.850		−0.062		−0.457	0.059	0.580	0.014	−0.032
2.15	0.050	(1.39)	(−3.04)	(−4.124)		(−3.82)		(−0.50)	(0.23)	(1.70)	(0.43)	(−4.09)
6. 1935–69	0.208	−3.503	−1.512	−0.424			−0.059	−1.368	0.485	1.455	0.064	−0.050
1.86	0.051	(−0.85)	(−1.94)	(−3.38)			(−1.73)	(−1.38)	(1.42)	(4.25)	(1.93)	(−4.87)

Note: All variables except T are in natural logarithms. The definitions of these variables are given in Table 2. The term Δ^*X denotes the linear operation $X-\hat{\beta}X_{-1}$. The value of β is estimated via the Cochrane-Orcutt iterative procedure (*CORC*). The term $\hat{\sigma}_e$ is defined in Section IIA1. The terms $\Delta^*\hat{P}a$ and $\Delta^*\hat{P}c|a$ in equations 1–5 are computed via a reduced form regression equation including: *C*(constant), Q/N_{-1}, Pa_{-1}, $Pc|a_{-1}$, $P^oe|c$, *L*, *A*, Y_p, *U*, T, $P^oe|c_{-1}$, L_{-1}, A_{-1}, Y_{p-1}, U_{-1}, $XPOL_1$, *XGOV*, *NW*, *N*. The terms $\Delta^*\hat{P}a$, $\Delta^*\hat{P}c|a$, and $\Delta^*\hat{P}XQ_1$ in equation 6 are computed via the same reduced form with $PXQ_1(P^oe|c)$ excluded.

lowest in absolute magnitude when the objective measure of $Pe|c$, PXQ_1, is treated in the regression analysis as if it were a perfectly forecast and strictly exogenous variable. The algebraic value of the elasticity associated with PXQ_1 is −0.039 with upper and lower 95 percent confidence limits (calculated from the normal distribution) of 0.008 and −0.086. The corresponding elasticities associated with the alternative measures of $Pe|c$, PXQ_{1-1}, PXQ_2, TXQ_1, $\hat{P}XQ_1$, and PDL_1

TABLE 4—MODIFIED FIRST DIFFERENCES OF MURDER RATES (IN NATURAL LOGARITHMS) REGRESSED AGAINST CORRESPONDING MODIFIED FIRST DIFFERENCES OF SELECTED VARIABLES SET II: ALTERNATIVE TIME PERIODS AND OTHER TESTS

($\hat{\beta}/S_{\hat{\beta}}$ in parentheses)

Effective Period *D.W.* Statistic	$\hat{\beta}$(*CORC*) $\hat{\sigma}_e$	*C* (Constant)	$\Delta^*\hat{P}^o a$	$\Delta^*\hat{P}^o c\|a$	$\Delta^*P^o e\|c$ Δ^*PXQ_{1-1}	 Δ^*TXQ_1	Δ^*L	Δ^*A	Δ^*Y_p	Δ^*U	War Years Dummy (1942–45)	Δ^*T
1. 1935–69[a]	0.059	−4.060	−1.247	−0.345	−0.066		−1.314	0.450	1.318	0.068		−0.046
1.80	0.044	(−1.00)	(−1.56)	(−3.07)	(−3.33)		(−1.49)	(2.20)	(4.81)	(2.60)		(−6.54)
2. 1937–69[a]	0.287	−2.568	−1.435	−0.474		−0.049	−1.388	0.526	1.289	0.063		−0.044
1.99	0.046	(−0.61)	(−1.87)	(−3.22)		(−2.31)	(−1.57)	(1.94)	(3.91)	(2.10)		(−4.96)
3. 1936–69[b]	—	−3.608	−1.385	−0.345		−0.064	−1.218	0.482	1.348	0.068		−0.047
1.49	0.046	(−1.03)	(−2.12)	(−3.25)		(−3.52)	(−1.40)	(2.13)	(4.94)	(2.59)		(−6.69)
4. 1935–69	0.061	−4.882	−1.172	−0.383	−0.069		−1.487	0.477	1.393	0.077	0.018	−0.048
1.84	0.046	(−1.32)	(−1.73)	(−3.20)	(−3.22)		(−1.61)	(1.89)	(4.30)	(1.95)	(0.31)	(−5.76)
5. 1937–69	0.250	−2.086	−1.634	−0.508		−0.055	−1.444	0.406	1.334	0.077	0.035	−0.045
2.08	0.048	(−0.51)	(−2.16)	(−2.83)		(−2.36)	(−1.51)	(1.23)	(3.73)	(1.80)	(0.50)	(−4.72)
6. 1941–69	−0.164	3.025	−1.744	−0.714	−0.074		−1.008	0.141	0.734	0.028		−0.036
2.21	0.048	(0.57)	(−2.21)	(−3.70)	(−3.70)		(−1.04)	(0.56)	(2.06)	(0.91)		(−4.40)
7. 1941–69	−0.029	3.752	−1.947	−0.723		−0.066	−0.962	0.152	0.771	0.0311		−0.036
2.13	0.048	(0.68)	(−2.38)	(−3.69)		(−3.34)	(−0.99)	(0.55)	(2.00)	(0.96)		(−4.13)
8. 1933–66	−0.001	−5.678	−0.564	−0.265	0.055		−2.111	0.283	0.922	0.036		−0.036
1.90	0.033	(−2.21)	(−1.10)	(−3.49)	(−3.72)		(−3.18)	(1.65)	(4.16)	(1.74)		(−6.30)
9. 1939–66	0.016	−2.601	−0.946	−0.360		−0.051	−1.766	0.212	0.780	0.027		−0.033
1.96	0.037	(−0.598)	(−1.38)	(−1.984)		(−3.23)	(−2.254)	(1.03)	(2.920)	(1.11)		(−4.99)

Note: same references as in Table 3 but the reduced form used to compute $\Delta^*\hat{P}^o a$ and $\Delta^*\hat{P}^o c|a$ does not include *N*.

[a] Same as equations 3 and 4 in Table 3 with the missing data pertaining to $XPOL_{-1}$ interpolated via a smoothing procedure.

[b] Same as equation 4 in Table 3 with $\hat{\rho}$ assumed to be zero (level regression).

vary between -0.049 and -0.068 with upper and lower 90 percent confidence limits ranging between -0.01 and -0.10. These results have been anticipated by the analysis of Section IIA2 where it was suggested that the regression coefficient associated with PXQ_1 is likely to be biased toward zero due to the effect of random forecasting errors. In addition, since the analysis of optimal social defense against murder suggests that an exogenous change in (Q/N) may change the socially optimal value of $Pe|c$ in the same direction, the coefficient associated with PXQ_1 may be biased toward a positive value because of a potentially positive correlation between (Q/N) and the unsystematic part of PXQ_1. This simultaneous equation bias is expected to be eliminated when the systematic part of PXQ_1 is estimated via the reduced form regression equation $(\hat{P}XQ_1)$. It is noteworthy that the estimated elasticities of $(Q/N)^0$ with respect to alternative measures of $Pe|c$ are found generally low in absolute magnitude. This, perhaps, is the principal reason why previous studies into the effect of capital punishment on murder using simple correlation techniques and rough measures of the conditional risk of execution have failed to identify a systematic association between murder and the risk of execution.

The regression results regarding the effects of P^0a, $P^0c|a$, and $P^0e|c$ constitute perhaps the strongest findings of the empirical investigation. Not only do the signs of the elasticities associated with these variables conform to the general theoretical expectations, but their ranking, too, is consistent with the predictions in Section I. Table 3 shows that the elasticities associated with P^0a range between -1.0 and -1.5, whereas the elasticities associated with $P^0c|a$ in the various regression equations range between -0.4 and -0.5. And, as indicated in the preceding paragraph, the elasticities associated with $P^0e|c$ are lowest in absolute magnitude. Consistent with predictions and evidence presented in Section IB regarding a negative association between $P^0e|c$ on the one hand and P^0a and $P^0c|a$ on the other, introduction of the latter variables in the regression equation is found to be particularly useful in isolating the (negative) deterrent impact of estimates of $P^0e|c$. Of similar importance is the introduction of the time trend T.

The estimated values of the elasticities associated with the unemployment rate U, labor force participation L, and permanent income Y_p in Table 3 are not inconsistent with the theoretical expectations discussed in Section IA. Of particular interest is that the effects of equal percentage changes in $P^0e|c$ and U are found to be nearly alike in absolute magnitude. Because murder is often a by-product of crimes involving material gains, the positive effect of U on $(Q/N)^0$ may be attributed in part to the effect of the reduction in legitimate earning opportunities on the incentive to commit such crimes. Indeed, preliminary time-series regression results show that the elasticities of robbery and burglary rates with respect to the unemployment rate are even larger in magnitude than the corresponding elasticities of the murder rate. These results conform more closely to theoretical expectations than do the results in a cross-state regression analysis (see the author (1973a)). The reason, presumably, is that due to their higher correlation with cyclical variations in the demand for labor, changes in U over time measure the variations in both involuntary unemployment and the duration of such unemployment more effectively than do variations in U across states at a given point in time. The estimated negative effect of variations in the labor force participation rate on the murder rate can be explained along similar lines. Theoretically, variations in L are likely to reflect opposing income and substitution effects of changes in market earning opportunities. However, with measures of both permanent income Y_p, and the rate of unemployment introduced in the regression equation as independent explanatory variables, changes in L may reflect a pure substitution effect of changes in legitimate earning opportunities on the incentive to commit crimes both against persons and property.[14] Finally, the positive association between Y_p and $(Q/N)^0$ need not imply a positive income elasticity of demand for hate and malice since changes in the *level* of the personal distribution of income may be strongly correlated with payoffs on crimes against property. If legitimate employment opportunities are effectively accounted for by U and by L, changes in Y_p may be highly correlated with similar changes in the incidence of crimes against property. Such a partial correlation is indeed observed across states and in a time-series regression analysis of crimes against property now in progress.

The positive effect of variations in the percentage of the population in the age group 15–24, A, on the murder rate is consistent with the cross-state evidence concerning the correlation between these variables. A possible explanation for this finding was already offered in Section IIA2. Additional analysis, not reported herein, indicated that the effect of the percentage of nonwhites in the population NW becomes statistically insignificant when the time trend T is introduced as an independent explanatory variable in the regression equation. Consequently, this variable is excluded from the regressions estimating the supply of murders function. This result stands in sharp contrast to the ostensibly positive effect of NW on the murder rate across states. I have argued elsewhere in this context that the apparently higher participation rate of nonwhites in all criminal activities may result largely from the relatively poor legitimate employment opportunities available to them (see the author (1973a)). Since, over time, variations in these opportunities may be effectively accounted for by the variations in U and L, the estimated independent effect of NW may indeed be nil. The negative partial effect of T on $(Q/N)^0$ reported in Tables 3 and 4 is not inconsistent with the predictions advanced in Section IIA2.

The regression results are found to be robust with respect to the functional form of the regression equation. In addition, estimating the regression equations by introducing the levels of the relevant variables rather than their modified first differences (that is, assuming no serial correlation in the error terms) artificially reduces the standard errors of the regression coefficients as would be expected on purely statistical grounds (see Table 4, equation (3)). The results are further insensitive as to the specific estimates of expenditures on police used in the reduced form regression equation. The data for this variable are not available for all the odd years between 1933 and 1951 and the missing statistics were interpolated either via a reduced form regression analysis (*XPOL*) or a simple smoothing procedure. The results are virtually identical (compare equations (1) and (2) in Table 4 with equations (3) and (4) in Table 3). The introduction of a dummy variable distinguishing the World War II years (1942–45) from other years in the sample has no discernible effect on the regression results, while the effect of the dummy variable itself appears to be statistically insignificant.

---

[14] A possible explanation for the significant negative association between labor force participation and particularly crimes against the person is that interpersonal frictions and social interactions leading to acts of malice occur mostly in the nonmarket or home sector rather than at work. An increase in the total time spent in the nonmarket sector (a reduction in L) might then generate a positive scale effect on the incidence of murder. This *ad hoc* hypothesis is nevertheless supported by FBI *UCR* evidence on the seasonal pattern of murder. This crime rate peaks twice a year: around the holiday season (December) and around the summer vacation season (July–August) in which relatively more time is spent out of work. It is also supported by evidence that the frequency of murders on weekends is significantly higher than on weekdays (see William Graves, p. 327).

Of more importance, the qualitative results reported in Table 3 are for the most part insensitive to changes in the specific interval of time investigated in the regression analysis, as indicated by the results reported in Table 4. However, the absolute magnitudes of some of the estimated elasticities, especially those associated with P^0a, $P^0c|a$, $P^0e|c$, U, and L do change when estimated from different subperiods. Finally, the time-series estimates of the supply-of-murders function appear quite consistent with independent estimates derived through a cross-state regression analysis using data from 1960. A detailed discussion of related issues is included in the author (1973b).

III. Some Implications

A. *The Apparent Effect of Capital Punishment: Deterrence or Incapacitation?*

It has already been hinted in the introduction to this paper that an apparent negative effect of execution on the murder rate may merely reflect the relative preventive or incapacitating impact of the death penalty which eliminates the possibility of recidivism on the part of those executed. An estimation of the differential preventive effect of execution relative to imprisonment for capital murder has been attempted in this study through an application of a general model of the preventive effect of imprisonment developed in the author (1973a). In this application of the model, execution is identified with an imprisonment term Te, which is equal to the life expectancy of an average offender imprisoned for murder. The differential preventive impact of execution is estimated by taking account of the alternative average sentence served by those imprisoned for capital murder Tm, the fractions of potential murders executed and imprisoned, and the rate of population growth.

Derivation of the expected partial elasticity of the murder rate with respect to the fraction of convicts executed, $\sigma_{P^0e|c}$, is omitted here for lack of space. I shall point out only that estimates of $\sigma_{P^0e|c}$ derived on the basis of the extremely unrealistic assumption that any potential murderer at large (outside prison) commits one murder each and every year and for values of Te and Tm estimated at 40, and between 10 and 16 years, respectively, vary between 0.020 and 0.037 (see the author (1973b)). These estimates, therefore, do not account for the full magnitudes of the absolute values of the elasticities of the murder rate with respect to estimates of the fraction of convicts executed that are reported in Tables 3 and 4. Moreover, according to the model of law enforcement involving only preventive effects, the partial elasticity of the murder rate with respect to the fraction of those apprehended for muder P^0a is expected to be identical to the corresponding elasticity with respect to the fraction of those apprehended and charged with murder who were convicted of this crime, $P^0c|a$. The reason, essentially, is that equal percentage changes in either P^0a or $P^0c|a$ have the same effect on the fractions of offenders who are incapacitated through incarceration or execution, and thus should have virtually equivalent preventive effects on the murder rate. This prediction is ostensibly at odds with the significant positive difference between empirical estimates of the murder rate with respect to P^0a and $P^0c|a$. In contrast, the latter findings are consistent with implications of the deterrent theory of law enforcement (see equation (3)). In light of these observations one cannot reject the hypothesis that punishment in general, and execution in particular, exert a unique deterrent effect on potential murderers.

B. *Tentative Estimates of the Tradeoff Between Executions and Murders*

The regression results concerning the partial elasticities of the reported murder rate with respect to various measures of the expected risk of execution given conviction in different subperiods $\hat{\alpha}_3$, can be restated in terms of expected tradeoffs between the execution of an offender and the lives of potential victims that might thereby be saved. For illustration, consider the regression coefficients associated with $\hat{P}XQ_1$ and $PXQ_{1_{-1}}$ in equations (6) and (3) of Table 3. These coefficients, -0.06 and -0.065, respectively, may be considered consistent estimates of the average elasticity of the national murder rate, $(Q/N)^0$, with respect to the objective conditional risk of execution, $P^0e|c=(E/C)^0$, over the period 1935–69. Evaluated at the mean values of murders and executions over that period, $\bar{Q}=8{,}965$ and $\bar{E}=75$, the marginal tradeoffs, $\Delta Q/\Delta E=\hat{\alpha}_3\bar{Q}/\bar{E}$, are found to be 7 and 8, respectively. Put differently, an additional execution per year over the period in question may have resulted, on average, in 7 or 8 fewer murders. The weakness inherent in these predicted magnitudes is that they may be subject to relatively large prediction errors. More reliable point estimates of the expected tradeoffs should be computed at the mean values of all the explanatory variables entering the regression equation (hence, also the mean value of the dependent variable) because the confidence interval of the predicted value of the dependent variable is there minimized. The mean values of the dependent variable and the explanatory variable used to calculate the value of $\hat{\alpha}_3$ in equation (3) of Table 3 are found to be nearly identical with the actual values of these two variables in 1966 and 1959, respectively. The corresponding values of murders and executions in these two years were $Q(1966)=10{,}920$ and $E(1959)=41$; the marginal tradeoffs between executions and murders based on the latter magnitudes and the elasticity $\hat{\alpha}_3=-0.065$ are found to be 1 to 17.

It should be emphasized that the expected tradeoffs computed in the preceding illustration mainly serve a methodological purpose since their validity is conditional upon that of the entire set of assumptions underlying the econometric investigation. In addition, it should be pointed out that the 90 percent confidence intervals of the elasticities used in the preceding illustrations vary approximately between 0 and -0.10 implying that the corresponding confidence intervals of the expected tradeoffs in the last illustration range between limits of 0 and 24. As the above illustrations indicate, however, although the estimated elasticities $\hat{\alpha}_3$ reported in Tables 3 and 4 are low in absolute magnitude, the tradeoffs between executions and murders implied by these elasticities are not negligible, especially when evaluated at relatively low levels of executions and relatively high levels of murder.[15]

Finally, it should be emphasized that the tradeoffs discussed in the preceding illustrations were based upon the partial elasticity of $(Q/N)^0$ with respect to measures of $P^0e|c$ and thus, implicitly, on the assumption that the values of all other variables affecting the murder rate are held constant as the probability of execution varies. In practice, however, the values of the endogenous variables Pa and $Pc|a$ may not be perfectly controllable. The theoretical analysis in Section IB suggests that exogenous shifts in the optimal values of $Pe|c$ may generate offsetting changes in the optimal values of Pa and $Pc|a$. Indeed, consistent estimates of the elasticities of the reported murder rates with respect to alternative measures of $P^0e|c$ that were derived through a reduced form regression analysis using as explanatory variables only the exogenous and predetermined variables included in the supply of offenses function and other structural equations (X_1 and X_2 in Table 2) are found to be generally lower than the elasticities reported in Table 3.[16] The actual tradeoffs between executions and murders thus depend partly upon the ability of law enforcement agencies to control simultaneously the values of all the parameters characterizing law enforcement activity.

---

[15] A decrease in the number of executions in 1960 from 44 to 2 (the actual number of executions in 1967), which implies a decline of 95 percent in the value of $Pe|c$ in that year, would have increased the murder rate that same year by about 6.2 percent from 0.05 to 0.053 per 1,000 population if the true value of α_3 were equal to 0.065. The implied increase in the actual number of murders in 1960 would have been from 9,000 to 9,558. For comparison, note that the actual murder rate in 1967 was 0.06 per 1,000 population and the number of murders was 12,100. The values of other explanatory variables associated with the supply of murders function were, of course, quite different in these two years. By this tentative and rough calculation, the decline in $Pe|c$ alone might have accounted for about 25 percent of the increase in the murder rate between 1960 and 1967.

IV. Conclusions

This paper has attempted to present a systematic analysis of the relation between capital punishment and the crime of murder. The analysis rests on the presumption that offenders respond to incentives. Not all those who commit murder may respond to incentives. But for the theory to be useful in explaining aggregate behavior, it is sufficient that at least some so behave.

Previous investigations, notably those by Sellin, have developed evidence used to unequivocally deny the existence of any deterrent or preventive effects of capital punishment. This evidence stems by and large from what amounts to informal tests of the sign of the simple correlation between the legal status of the death penalty and the murder rate across states and over time in a few states. Studies performing these tests have not considered systematically the actual enforcement of the death penalty, which may be a far more important factor affecting offenders' behavior than the legal status of the penalty. Moreover, these studies have generally ignored other parameters characterizing law enforcement activity against murder, such as the probability of apprehension and the conditional probability of conviction, which appear to be systematically related to the probability of punishment by execution. In addition, the direction of the causal relationship between the rate of murder and the probabilities of apprehension, conviction, and execution is not obvious, since a high murder rate may generate an upward adjustment in the levels of these probabilities in accordance with optimal law enforcement. Thus the sign of the simple correlation between the murder rate and the legal status, or even the effective use of capital punishment, cannot provide conclusive evidence for or against the existence of a deterrent effect.

The basic strategy I have attempted to follow in formulating an adequate analytic procedure has been to develop a simple economic model of murder and defense against murder, to derive on the basis of this model a set of specific behavioral implications that could be tested against available data, and, accordingly, to test those implications statistically. The theoretical analysis provided sharp predictions concerning the signs and the relative magnitudes of the elasticities of the murder rate with respect to the probability of apprehension and the conditional probabilities of conviction and execution for murder. It suggested also the existence of a systematic relation between employment and earning opportunities and the frequency of murder and other related crimes. Although in principle the negative effect of capital punishment on the incentive to commit murder may be partly offset, for example, by an added incentive to eliminate witnesses, the results of the empirical investigation are not inconsistent with the hypothesis that, on balance, capital punishment reduces the murder rate. But even more significant is the finding that the ranking of the elasticities of the murder rate with respect to Pa, $Pc|a$, and $Pe|c$ conforms to the specific theoretical predictions. The murder rate is also found negatively related to the labor force participation rate and positively to the rate of unemployment. None of these results is compatible with a hypothesis that offenders do not respond to incentives. In particular, the results concerning the effects of the estimates of the probabilities of apprehension, conviction, and execution are not consistent with the hypothesis that execution or imprisonment decrease the rate of murder only by incapacitating or preventing apprehended offenders from committing further crimes.

---

[16] The elasticities associated with PXQ_1, $PXQ_{1_{-1}}$, TXQ_1, and PDL_1 in this modified reduced form regression analysis relating to the period 1934–69 are found equal to −0.0269 (−0.83), −0.0672 (−2.29), −0.0414 (−1.99), and −0.052 (−5.81), respectively, where the numbers in parentheses are the ratios of the coefficients to their standard errors.

These observations do not imply that the empirical investigation has proved the existence of the deterrent or preventive effect of capital punishment. The results may be biased by the absence of data on the severity of alternative punishments for murder, by the use of national rather than state statistics, and by other imperfections. At the same time it is not obvious whether the net effect of all these shortcomings necessarily exaggerates the regression results in favor of the theorized results. In view of the new evidence presented here, one cannot reject the hypothesis that law enforcement activities in general and executions in particular do exert a deterrent effect on acts of murder. Strong inferences to the contrary drawn from earlier investigations appear to have been premature.

Even if one accepts the results concerning the partial effect of the conditional probability of execution on the murder rate as valid, these results do not imply that capital punishment is necessarily a desirable form of punishment. Specifically, whether the current level of application of capital punishment is optimal cannot be determined independently of the question of whether the levels of alternative punishments for murder are optimal. For example, one could argue on the basis of the model developed in Section IA that if the severity of punishments by means other than execution had been greater in recent years, the apparent elasticity of the murder rate with respect to the conditional probability of punishment by execution would have been lower, thereby making capital punishment ostensibly less efficient in deterring or preventing murders. Again, this observation need not imply that the effective period of incarceration imposed on convicted capital offenders should be raised. Given the validity of the analysis pursued above, incarceration or execution are not exhaustive alternatives for effectively defending against murders.[17] Indeed, these conventional punishments may be considered imperfect means of deterrence relative to monetary fines and other related compensations because the high "price" they exact from convicted offenders is not transferrable to the rest of society. Moreover, the results of the empirical investigation indicate that the rate of murder and other related crimes may also be reduced through increased employment and earning opportunities. The range of effective methods for defense against murder thus extends beyond conventional means of law enforcement and crime prevention. There is no unambiguous method for determining whether capital punishment should be utilized as a legal means of punishment without considering at the same time the optimal values of all other choice variables that can affect the level of capital crimes.

---

[17] Ironically, the argument that capital punishment should be abolished because it has no deterrent effect on offenders might serve to justify the use of capital punishment as an ultimate means of prevention of crime, since the risk of recidivism that cannot be deterred by the threat of punishment is not eliminated entirely even inside prison walls. In contrast, since the results of this investigation support the notion that execution exerts a pure deterrent effect on offenders, they can be used to suggest that other punishments, even those which do not have any preventive effect, can in principle serve as substitutes.

REFERENCES

C. B. Beccaria, *An Essay on Crimes and Punishments,* London 1767, originally published in 1764.

H. A. Bedau, *The Death Penalty in America,* Garden City 1967.

G. S. Becker, "Crime and Punishment: An Economic Approach," *J. Polit. Econ.*, Mar./Apr. 1968, *78,* 169–217.

———, "A Theory of Social Interactions," *J. Polit. Econ.*, Nov./Dec. 1974, *82,* 1063–93.

I. Ehrlich, "Participation in Illegitimate Activities: An Economic Analysis," unpublished doctoral dissertation, Columbia Univ. 1970.

———, (1973a) "Participation in Illegitimate Activities: A Theoretical and Empirical Investigation," *J. Polit. Econ.*, May/June, 1973, *81,* 521–65.

———, (1973b) "The Deterrent Effect of Capital Punishment: A Question of Life and Death," Nat. Bur. Econ. Res. working pap. series no. 18, 1973.

R. C. Fair, "The Estimation of Simultaneous Equation Models with Lagged Endogenous Variables and First Order Serially Correlated Errors," *Econometrica,* May 1970, *38,* 507–16.

W. F. Graves, "The Deterrent Effect of Capital Punishment in California," in H. A. Bedau, ed., *The Death Penalty in America,* Garden City 1967.

H. M. Hochman and J. D. Rodgers, "Pareto Optimal Redistribution," *Amer. Econ. Rev.*, Sept. 1969, *59,* 542–57.

T. Sellin, *The Death Penalty,* Philadelphia 1959.

———, *Capital Punishment,* New York 1967.

President's Commission on Law Enforcement and Administration of Justice *(PCL), Crime and Its Impact—An Assessment,* Task Force Reports, Washington 1967.

U.S. Department of Justice, Bureau of Prisons, *National Prisons Statistics Bulletin (NPS),* various numbers; published prior to 1950 by Bureau of Census, Washington.

———, Federal Bureau of Investigation, *Uniform Crime Report (UCR),* Washington, various years.

U.S. Office of Business Economics, *The National Income and Product Accounts of the United States, 1929–65, Statistical Tables, Surv. Curr. Bus.*, supp., Washington 1966.

———, *Business Statistics, Surv. Curr. Bus.*, Washington, various years.

China as a Less Developed Economy

By Lloyd G. Reynolds*

The body of development theory which has grown up over the past generation assumes private ownership in both agriculture and industry. The public sector is omitted or assigned only a limited role in the economy. This is true, for example, of the well-known models of John Fei and Gustav Ranis, Dale Jorgenson, Shigeru Ishikawa (1967), and Allen Kelley, Jeffrey Williamson, and Russell Cheatham. There is as yet no theory of socialist development from a low income level.

This note makes no effort to construct such a theory. It has the modest purpose of underlining some aspects of China's development strategy which may be suggestive for model builders. The discussion relates mainly to Chinese experience since 1958. Up to that point Chinese development, which was strongly Soviet influenced as well as Soviet aided, showed the usual socialist emphasis on rapid industialization and a predominant allocation of investment resources to heavy industry. This phase is well documented in the studies by Alexander Eckstein, Walter Galenson, and Ta-Chung Liu, and by Nai-Ruenn Chen and Galenson.

From the "great leap" program initiated in the winter of 1957–58, and from the Chinese-Soviet rift which followed soon thereafter, there ensued a sequence of institutional changes and policy shifts which is still continuing. But while the economy has in no sense come to rest, and while statistical estimates from 1958 onward are conjectural, one can discern the outlines of a development strategy distinctly different from the Soviet pattern, and different also from that of other large, densely populated Asian countries such as India, Indonesia, and Pakistan.

I. The Planners' Objective Function

The development literature of the 1950's and 1960's tended to assume that development consisted in a more rapid growth of *GNP* per capita. But it gradually became apparent that even in countries where output growth was satisfactory, employment growth was considerably slower. The backlog of surplus labor was not being reduced or was even increasing. In many cases, too, additional output was distributed so as to accentuate preexisting inequalities of income. The recent tendency, therefore, has been to downgrade *GNP* increase alone as a success indicator, and to erect instead a trinity of objectives: full employment, a high growth rate of *GNP* per capita, and a more equitable distribution of income.

The objectives of Chinese leaders are in line with this current thinking—indeed, the Chinese were there first. Full mobilization of manpower (and, importantly, of womanpower) was an objective from the beginning, and had been largely accomplished by the early 1950's. The Chinese leadership has also pursued an incomes policy aimed at 1) elimination of poverty, in the sense of establishing a minimum standard of living to which every citizen is entitled; and 2) gradual reduction of income inequalities above the minimum, by methods which will be noted below.

Thus far the three objectives do not

---

* Yale University. Some of the observations in this note stem from a recent visit to China with four economist colleagues, which included extended discussions with Chinese economists and management officials. I have had access also to a number of other travel reports, in addition to published English language sources on the Chinese economy.

seem to have been seriously in conflict. The additional employment which has been created, most of which is necessarily in the countryside, has had a positive marginal product and has thus aided the growth of output. Income equalization has not been pushed to the point where it interferes seriously with work incentives. On the contrary, the fact that many who previously suffered from disease and malnutrition are now healthier and better fed must have increased productive capacity as well as welfare.

In addition to economic objectives whose accomplishment is measurable in principle, several less tangible objectives have an important influence on economic decisions. These objectives, which have been described by John Gurley, Joan Robinson, and others, include:

(i) *Deemphasis of material incentives.* Enterprise managers are paid little more than skilled workers. Profits go entirely to the state, with no bonuses to workers or managers for superior performance. The object is to establish a motivational system in which people will work effectively from social conscience, reinforced by group pressure from fellow workers, rather than for monetary rewards.

(ii) *Antibureaucratization.* In any industrial society, capitalist or socialist, there is a tendency for government officials, industrial managers, and professional experts to crystallize into an élite—Djilas' "ruling class," or Galbraith's "technostructure." Chinese leaders are determined to maintain social fluidity to prevent the emergence of marked distinctions, either of prestige or income. This underlies, for example, the new policy of requiring high school graduates to work for at least two years before they can be considered for university admission, and of giving preference in admission to children of workers and farmers. This is intended to prevent hereditary transmission of status through the educational system.

(iii) *Mass participation in decision making.* This is evidenced in worker membership in the "Revolutionary Committees" which now govern all organizations in China; in the rapid promotion of factory workers to leadership positions during and since the Cultural Revolution; in the establishment of "three-in-one teams," including managers, technicians, and workers, as the responsible groups for industrial research and development. The opposite side of the coin is the downgrading of the expert, the specialist, the technician, who is urged to "learn from the workers," and to spend frequent periods of time in manual labor.

These cardinal points of Maoist doctrine raise difficult questions, to which there are presently no clear answers. How widely are they pursued as well as advocated? Are they feasible objectives over an extended period of time? To what extent does pursuit of these objectives conflict with efficient occupational division of labor, scientific and technical progress, and increase of productivity? Chinese leaders would probably deny the existence of any such conflict. But even if they were persuaded that a tradeoff exists, they would doubtless accept some slowdown of material progress as a reasonable price for better performance in social organization and human relations.

We turn now to China's treatment of several problems which are endemic in the *LDC*s, and particularly in the densely populated countries of South and East Asia: increasing the rate of capital accumulation; producing and transferring adequate food supplies for the urban population; reducing the rate of population growth; preventing premature migration to the cities and the associated rise of urban unemployment; providing adequate employment opportunities in the countryside; and coping with the tendency toward growing income inequality which often marks the early decades of modern economic growth.

II. Devices for Capital Accumulation

While capital is not the sole requisite for growth as it seemed in some early development models, it remains an important requirement. In the older industrial countries, gross capital formation as a percentage of *GNP* rose gradually during the early decades of accelerated growth; and the more successful *LDC*s today are repeating this experience.

The unusual feature of the Chinese case is the suddenness with which the capital formation rate was raised from a low plateau to a much higher plateau. Kung-Chia Yeh, in his contribution to the Eckstein, Galenson, and Liu volume, estimates that the ratio of gross domestic capital formation to *GNP*, measured in constant 1952 prices, averaged 7.4 percent from 1931–36 and 24.0 percent from 1952–57. Shigeru Ishikawa's (1965) estimates for the 1950's are also in the 20–25 percent range.

A major factor in this early period was the redistribution of income associated with the land reform of 1949–52. A substantial share of national income, variously estimated at 15 to 30 percent, had previously gone to landlords, rich gentry, money-lenders, merchants, and others. Land reform redistributed this income, on the first round, among the peasant population. Government then intervened, through tax and price instruments, to capture much of this income and redirect it toward capital formation. There is a parallel here with Japanese government strategy in the early Meiji period, when diversion of income from the previous landed nobility into government channels helped to raise the national savings rate.

While comparable data are not available for more recent years, it seems unlikely that the capital formation rate today is *lower* than in 1952–57. Dwight Perkins estimates that from 1957 to 1972 output of producers' goods rose at an average annual rate of 10 percent per year, compared with 8 percent for all industrial output; and that the share of producers' goods in *GNP* was twice as high in 1972 as in 1957. Granted that part of this increased output went for military purposes, and that the output of heavy industry does not measure construction and other parts of capital formation, the data suggest a continued high capital formation rate.

Within the limits of consumer tolerance, a socialist economy can determine the capital formation rate through control of financial flows as well as through physical resource allocation. There is a price-cost margin for manufactured goods, the allocation of which between taxes and profits is a matter of discretion. In Chinese practice, as distinct from Soviet practice, excise taxes are relatively low and profits correspondingly high; and all profit is remitted to government rather than being shared with the enterprise (though enterprise control over depreciation allowances doubtless permits some net investment under the guise of reinvestment). The price-cost margin varies with the product, being low for agricultural inputs, machinery, and essential consumer goods, and higher for less essential consumer items. Thus, although China has no income tax, there is an element of progressive income taxation in the pricing system.

Another distinctive feature of the Chinese system is the relatively great importance of lower levels of government. The great bulk of industrial activity, and all agricultural activity, is directed by the provinces, counties, and communes. The provinces are accordingly the key revenue-collecting agency. Only about 20 percent of central government revenue is collected directly from national industries and other sources. The remainder comes through transfers from the provinces, which are varied according to the province's ability to pay. This might be termed "revenue sharing from below." Major investment schemes are included in the central government budget, but the bulk of investment is planned and financed via provincial and county budgets.

Much investment, indeed, is carried out at the lowest level of the rural commune. Each production team contributes 3 to 5 percent of its gross income to an "accumulation fund," which finances equipment purchases and other agricultural improvements by the team. The production brigade and the commune draw revenue from machinery leases and other services to the production teams, and also from the profits of their industrial activities. The profits of rural industrial enterprises, unlike the profits of urban industries, are not remitted to the state. The funds remain in the commune and are available for expansion of industrial activities or for agricultural investment. Commune officials report that their present scale of industrial production has been built up largely from reinvested earnings rather than through investment allocations from the state.

III. Generating an Agricultural Surplus

In socialist economies agriculture is typically a lagging sector. It is not clear to what extent this is due to inherent inefficiency of collective or state farming compared with the private "family farm," and how far it stems from antiagriculture bias on the part of policy makers, reflected in underinvestment in agriculture and in pricing and procurement policies which provide little incentive for output increases.

Chinese economic policy has been distinctly more favorable to agriculture than has traditional Soviet policy. Agriculture was not tightly squeezed even during the industrialization drive of the 1950's. While investment allocations to agriculture from the state budget appear small, much of agriculture's saving was left for reinvestment within that sector. The drop in agricultural output during the "great leap" campaign stimulated even greater attention to agricultural output from 1960 onwards.

Much effort has been put into irrigation, pumping, and other water-control systems, permitting both an increase in multiple cropping and heavier fertilizer applications. Production of fertilizer and farm machinery has risen rapidly in percentage terms, though from a low base. Use of modern inputs is encouraged by pricing them close to cost of production. Government procurement prices for grain and other major crops have been raised moderately since 1960, while prices of manufactures have tended downward, so that agriculture's terms of trade have been improving. Agricultural taxation is moderate, taking currently about 6 percent of gross agricultural output. This percentage has declined over time because assessments, which are based on "normal yields," lag behind the increase in output capacity. For non-key products (vegetables, fruits, poultry, eggs, pigs, milk, forestry products, fish), the communes have a high degree of autonomy. Delivery contracts are made directly with municipal trading authorities in nearby cities, at locally negotiated prices.

This package of incentives has stimulated a considerable increase in farm output. Perkins estimates that over the period 1957–72, agricultural output rose at an average annual rate of 3 percent. With population growing at something like 2 percent,[1] this provides a modest margin for dietary improvement.

The 80 percent of the people living in the countryside have to feed the 20 percent in towns and cities. This food transfer occurs in four main ways: (a) the agricultural tax, which is usually paid in grain; (b) quota sales of grain, cotton, sugar, and other key crops to the state procurement organization, at a uniform national price. Quotas appear to be determined as follows: each production team estimates its next year's production of the crop in question, based on an average of recent years. It then (in the case of grain) deducts the basic grain ration for the population of the team. The remainder is the amount which it can undertake to sell. (c) Above-quota sales, which are always permissible if the crop year is unusually good or if there has been a misjudgment of production capacity. For grain, though not for cotton and other cash crops, above-quota sales command a 30 percent price premium. (d) The direct commune-city contracts already mentioned govern most transfers of fruits, vegetables, and livestock products.

Ability to produce and transfer sufficient grain to the cities has enabled the city price of food grains to remain unchanged since the early 1950's. This fact, plus similar stability in the price of cloth and other necessities, removes upward pressure on the wage structure. There have been wage increases, particularly for the lowest paid workers, but these increases have been moderate and entirely discretionary. Wage stability combined with productivity improvement means that unit production costs of manufactured goods are typically falling over time, ensuring profit margins which contribute to capital accumulation and which also permit occasional price reductions, especially for the still relatively expensive consumer durables (bicycles, radios, watches, sewing machines).

The wage-price stability which is one of the unusual features of the Chinese economy thus rests essentially on the price of grain. The importance of preventing any upward pressure on this key price is doubtless one reason, though by no means the only reason, for maintenance of the rationing system despite apparently adequate grain supplies.

If domestic food supplies are adequate, why is China still importing substantial amounts of grain? One can only surmise. It is said that grain inventories are being increased, partly for defense reasons, at all levels from the farm household to the central government. Further, China is a very large country, with food-surplus and food-deficit provinces, and a limited transportation system. It may be efficient to supplement supplies in some of the coastal areas through imports rather than internal shipments. The imports also give China more elbowroom to export food to other less developed countries in the interest of political relations.

While the Chinese are keeping their heads above water on the crucial food front, serious problems remain in the agricultural sector. A central problem is how to generate and disseminate improved agricultural technology. The soil- and climate-specific characteristics of agricultural production make technology transfer from advanced economies more difficult than in industry. Much research and experimentation is usually required to adapt imported seed strains and cultivation methods to local conditions, as well as to generate indigenous improvements. China has a small supply of agricultural scientists. Nor is the present political climate—suspicion of intellectuals, downgrading of professional expertise and specialization, disorganization of the university system—conducive to enlargement of this supply.

---

[1] Neither the size of the population nor its rate of increase is known with any precision. Estimates for the 1950's, reported in Eckstein, Galenson, and Liu, are in the neighborhood of 2 percent. But today's rate of natural increase may be somewhat lower because of intensified efforts at family planning.

A second problem is how the small supply of improved seeds, chemical fertilizer, and other modern inputs should be allocated. Heavy allocations to farm areas which are already relatively productive might do most to raise national output; but it would also widen the income gap between more and less productive farms. On the other hand, allocations to the low-income farms on equity grounds might fail to maximize total output. Little is known about the direction of allocations at present. At any event, the dilemma remains.

IV. Restraining Population Growth

Population increase is the other side of the food problem. Crudely, reducing the population growth rate by 1 percent is as helpful as raising the growth of agricultural output by 1 percent, and in principle is much less expensive. Here China starts with several advantages over most other *LDC*s: an unambiguous commitment to population control by the political leadership; a system of primary education which is already universal in the cities and rapidly becoming so in the countryside; a widespread network of medical clinics and dispensaries, staffed largely by paramedical personnel to complement the small corps of fully trained doctors; and strong social discipline which is perhaps traditionally Chinese, but which is now reinforced by neighborhood associations and Communist Party units.

The most important single device is late marriage. Women are strongly urged not to marry before 25, and men not before 30. After marriage, the two-child family is held up as ideal. Contraceptive information and supplies are available without charge at the dispensaries. Carl Djerassi estimates that about half the urban women of reproductive age use oral contraceptives, and that the 1972 production rate was sufficient to supply 20 million women per year. After a couple has had two children, they are reportedly urged by the neighborhood association to stop at that point. The fact that women are encouraged to work equally with men, and that jobs are always available, provides an additional incentive to family planning. It is obviously easier for a woman to hold a full-time job with two children than with six.

Aggregative data on fertility rates are not available. But one can do amateur demography by inquiring into such things as school enrollments in each primary grade in a commune, or number of babies born in recent years in a housing development of known population. There is good reason to think that city birth rates are falling quite rapidly. Progress in the countryside has admittedly been slower; but the effort is continuing, and the announced objective of reducing the population growth rate to 1 percent (that is, to "developed country" levels) by 2000 A.D. does not seem unrealistic.

V. The Level of Employment

In China, as elsewhere, there are strong incentives to exchange farm work for urban employment. Work in the city is lighter, earnings are higher, educational and cultural amenities are greater. In most *LDC*s the result has been cityward migration in excess of urban labor requirements, a rise in unemployment, growth of "shanty towns" on the edge of the city, increased congestion and pressure on municipal services, and frequently political unrest and violence.

The Chinese use several policy instruments to avert these developments. Perhaps most important is industrial location policy. The announced objective is that the major eastern cities should not increase in size. In general, no new industry will be located in these cities, and there is strong effort to resist employment increases in older industries. New industries in the eastern provinces are located in suburban areas well away from the city center, and with their own housing and shopping facilities. In addition, more than half of industrial investment is now being directed toward the middle and far western provinces.

Other policy instruments include: 1) control of hiring. Enterprises may not recruit workers directly, but must apply to a municipal employment bureau. If an enterprise asks for additional labor, the first response is to urge it to get more output from its present labor force. Second, the bureau tries to make sure that the potential labor supply from the existing urban population is being fully utilized. Only as a last resort will new workers be accepted from the countryside. 2) Control of housing allocations. A would-be migrant from the country who applies for an apartment in the city is asked, "Do you have a job?" If the answer is no, he is back to square one. 3) The food and clothing rationing system. Every citizen has a "ration book location." To change this requires lengthy paperwork, and permission is given only for good cause. This network of controls makes it almost impossible to be unemployed in the city.

Not only are there restraints on rural-urban migration, but the flow of labor is sometimes in the other direction. There have been years recently in which the number of new jobs available in Canton or Shanghai was insufficient to accommodate all the school-leavers. In this event, the labor surplus is simply dispatched to the country for farm work.

The effectiveness of migration policy is suggested by the fact that the proportion of Chinese population living in rural areas has been holding steady at about 80 percent, instead of falling as one might expect in the course of economic growth.

But does the damming up of population in the countryside not mean simply that full employment in the city is achieved at the cost of growing rural underemployment? In many less developed countries, such underemployment is a chronic and growing problem. The Chinese assert that, on the contrary, their rural labor force is fully employed and that labor is scarce. If the Chinese claim is correct, the number of man-hours being worked in the rural sector today must be ***much*** larger than in 1949; and the sources of this increase need to be explored.

It appears that there has indeed been a major increase in rural employment since 1949. The main components of this increase seem to have been:

1) There has been a marked increase in the labor time devoted to infrastructure activities: roads, water control and irrigation systems, land terracing, reforestation, building of schools, hospitals, public buildings, and housing. These activities are organized mainly within a commune; but large-scale projects may be planned at higher levels and draw labor from several communes. The work is done mainly in the off-season, and helps to fill the valleys in the seasonal cycle of agricultural activity.

2) These infrastructure activities, and especially improved water control, have enabled a considerable shift of acreage from single to double cropping. They also make possible greater use of fertilizer and more sophisticated input "packages," which substantially increase crop yields. All this requires more labor.

3) Particularly in areas near city markets, there has been a diversification of production into fruits and vegetables,

poultry, eggs, milk, and meat. These activities are more labor intensive than grain growing.

4) The commune is an industrial as well as an agricultural producing unit. While the romantic vision of iron smelters in every village has understandably abated, there has been substantial development of industries which can operate efficiently on a relatively small scale. These rural industries produce partly for farm use—tractor and implement repair shops, fertilizer plants, brick kilns, cement plants, power generating units, water pumps. But they also turn out consumer goods for off-farm sale and even for export. At seasonal peaks of harvest activity, most of these industrial workers go into the fields; but their *main* activity is factory work.

5) The expansion of education, health and other public services in the rural areas absorbs substantial amounts of labor. Education is very labor-intensive, using the time of students as well as teachers. Medical and paramedical personnel in a typical commune would run into the dozens. Administrative staffs, which perform local government as well as production-management functions, are also substantial, though all except a few of the top commune officials also engage regularly in farm labor.

All in all, it appears that the claims of full employment in the rural sector are valid. Some of this employment, to be sure, has a low marginal yield per man-hour; for example, creating new land with great effort by carrying baskets of earth from some other area. The available labor time is apparently regarded as a fixed factor, and any small addition it can make to output is considered worthwhile. China is thus still a "labor surplus economy" in the sense that there are many rural workers whose marginal product is below their consumption. But until there is a demand for these workers in urban industries, their use in low-productivity rural activities is sound social policy.

VI. The Level and Distribution of Income

In considering personal income, we start from the usual distinction between its average level and its distribution among households. Following the revolution, the mass of the population experienced a once-and-for-all gain through income redistribution. Land rents, other property income, and such income sources as speculation, gambling, bribery, and extortion were abolished. Granted that the redistibution process may have affected the amount of income, it seems clear that part of the consumption of the former rich became available to lower-income groups. In the cities, a visible form of transfer was the subdivision of large private homes into multiple-family dwellings. In the countryside, income formerly paid to the landlord was available for increased consumption as well as capital formation and tax payments.

Since this initial redistribution, the average level of per capita consumption has continued to rise through the growth of national output. Perkins estimates that from 1957 to 1972, the output of manufactured consumer goods rose at about 4 percent per year. The estimate of 3 percent per year for agricultural output has already been noted. The rate of increase in educational services, health services, and other public goods has been substantial. Assuming that population has grown at no more than 2 percent per year, one can reasonably conclude that per capita consumption has been rising at 1 to 2 percent a year (probably closer to the latter figure).

Observation on a visit to China suggests that this improvement takes such concrete forms as: greater variety of diet, including a gradual increase in consumption of animal proteins; a gradual increase in the percentage of the population living in newly constructed housing; and a growing surplus of family income over basic expenditure, which can be used for unrationed clothing items, travel, and such consumer durables as furniture, bicycles, sewing machines, radios, and wrist watches.

The average level of living remains very low by Western standards; but it is not as low as suggested by *GNP* estimates converted at official exchange rates. These indicate a Chinese *GNP* per capita in the neighborhood of $150, which would imply that the Chinese have long since starved to death. The difficulty is that the official exchange rate differs widely from the purchasing power parity of the two currencies. The Yuan is grossly undervalued; and the discrepancy has been increasing as prices in the United States continue to rise while the Chinese price level remains completely stable—indeed, slightly falling.

One member of our visiting group collected a good deal of retail price information in Shanghai, Tientsin, and Peking; and because he spoke fluent Chinese, he was able also to collect some family budgets for city workers. One interesting comparison involved taking a Chinese market basket and inquiring how much a relatively well-off workers' family in Shanghai would spend on each item of consumption. This totaled 150 Yuan per month, or about $75 at official exchange rate. After returning to the United States, he estimated how much it would cost for a family in Ann Arbor, Michigan, to live *in the same way* (i.e., eating mainly cereals with very little animal products, using public transportation rather than a private car, living in a two-room apartment, etc.) The total came to about $400 per month, or $4,800 per year.

The conceptual difficulties of this kind of comparison are well known; but crude as it is, it suggests the wisdom of avoiding simply exchange rate conversions. While Chinese living standards are low, they are not abysmal.

Turning to distribution, the degree of income equalization thus far achieved in China should not be exaggerated. There are wide differences in per capita income among the rural communes, related to differences in soil fertility, water availability, climate, distance to urban markets, and other factors. Perkins reports a 1965 survey of thirteen communes which revealed that income per worker in the richest commune was 3.4 times as high as in the poorest. The average level of rural incomes is substantially below that of urban incomes. Within the urban sector, monthly incomes range from 30 to 40 Yuan a month for an unskilled factory worker to 110 Yuan for a skilled worker and as much as 400 Yuan for a top civil servant or university professor.

The Chinese have succeeded, however, in cutting off the extreme "tails" of the income distribution. There are no longer any rich people—no large houses, private automobiles, staffs of servants, expensive foreign vacations, conspicuous display of luxury. Nor are there any beggars or destitute people. No one is allowed to fall below a minimum level of living. Abolition of poverty in this sense, in a still poor and backward economy, is surely a major accomplishment.

The devices used for poverty elimination include: 1) free education and virtually free medical care; 2) a pricing system which discriminates in favor of basic food and clothing items, apartment rents, public utility services, public transportation, and other essential items, and which is reinforced by rationing of food grains, vegetable oils, cloth, and housing space; 3) an incomes policy which involves raising the lowest wage rates from time to time while leaving higher rates unchanged; 4) high demand for labor plus elimination of sex discrimination in employment, which means that city families are usually two-income families, whose combined earnings are sufficient to buy the minimum bundle of necessities; 5) special provision for the aged, including a retirement pension equal to 70 percent of previous earnings plus old people's homes for those unable to live with children. The traditional Chinese family system, under which young people live with the family and contribute to its income until marriage, while old people also remain within the family unit, is also helpful in solving the familiar American problem of the isolated, impoverished individual.

While inequalities above the minimum remain substantial, the thrust of government policy is toward whittling down these inequalities gradually over the years. The continuing increase in farm procurement prices is intended to reduce the rural urban income gap. Wage adjustments in the city typically involve raising rates for the lowest two or three labor grades without raising the top of the structure, thus compressing occupational differentials. Expansion of the supply of subsidized necessities works in the direction of equalizing *real* incomes.

VII. Commentary

A subject of perennial interest to development economists is which of the less developed countries have embarked on sustained economic growth. The usual criterion is a statistically reliable growth rate of income per capita, extending over several decades so that it cannot be dismissed as cyclical or transient. China seems clearly to meet this criterion. Income per capita has grown at a modest rate since 1949. Further, the organization of the economy and the strength of political leadership give reasonable assurance that growth will continue in the future. Greatest uncertainty attaches to the outlook for agriculture. But even here the fact that China has scarcely begun to tap the potential of modern biological-chemical innovations in agriculture provides some grounds for optimism.

There has been progress also in other dimensions of development. The crucial population problem has been faced squarely and is in course of being brought under control. Employment opportunities have been enlarged and sex discrimination in employment much reduced. Dire poverty has been eliminated, and there is a continuing reduction of income inequality above the minimum.

It would be overly facile, however, to conclude that Chinese experience provides lessons readily transferable to other *LDC*s. Economic policies cannot be wrenched out of the political and social context from which they have emerged. China did, after all, have a full-scale revolution. The willingness of the population to accept sustained hard work plus austerity in consumption, while it may in part be traditionally Chinese, depends also on a quasi-religious devotion to Maoist doctrine.

A more modest and tenable conclusion is that any *LDC* can profit from close examination of Chinese experience. Especially for the countries of South and Southeast Asia, this experience is more relevant than that of the Soviet Union or the East European countries.

The other experience which *LDC*s are usually urged to study is that of Japan. So it is worth noting that despite differences in the institutional context, there are similarities between recent Chinese development and Japanese development between, say, 1880 and 1905. In both cases one observes: 1) a marked emphasis on agricultural development, leading to output increases of about 1 percent per capita per year; 2) a marked development of rural-based industry, organized in relatively small and labor-intensive workshops, and employing mainly farm family members; 3) rapid development of transport and communications; 4) establishment of universal primary education; 5) establishment of numerous heavy industrial enterprises under government auspices, stimulated partly by military requirements.

There are also obvious differences in such matters as rate of labor transfer from agriculture, degree of involvement in foreign trade, and trends in income distribution. But China-Japan comparisons should evenutally tell us more about whether economies starting with similar factor endowments and income levels tend to follow a similar development path despite institutional differences.

REFERENCES

N-R. Chen and W. Galenson, *The Chinese Economy Under Communism*, Chicago 1969.

C. Djerassi, "Birth Control in the People's Republic of China," in *Bulletin of the American Academy of Arts and Sciences*, May 1974, *27*, 36–43.

A. Donnithorne, "China's Cellular Economy: Some Economic Trends Since the Cultural Revolution," *China Quarterly*, Oct.-Dec. 1972, *52*, 1–15.

A. Eckstein, W. Galenson, and T-C. Liu, *Economic Trends in Communist China*, Chicago 1968.

J. C. H. Fei and G. Ranis, *Development of the Labor Surplus Economy: Theory and Policy*, Homewood 1964.

J. G. Gurley, "Capitalist and Maoist Economic Development," in E. Friedman, ed., *America's Asia*, New York 1971.

S. Ishikawa, *National Income and Capital Formation in Mainland China*, Tokyo 1965.

———, *Economic Development in Asian Perspective*, Tokyo 1967.

D. Jorgensen, "Development of the Dual Economy," *Econ. J.*, June 1961, *71*, 309–34.

A. C. Kelley, J. G. Williamson, and R. I. Cheetham, *Dualistic Economic Development: Theory and History*, Chicago 1972.

D. H. Perkins, "Looking Inside China: An Economic Reappraisal," *Problems of Communism*, May-June 1973, *22*, 1–13.

J. Robinson, *Economic Management: China, 1972*, Modern China series, no. 4, London 1973.

S. Swamy, "Economic Growth in China and India, 1952–1970: A Comparative Appraisal," *Econ. Develop. Cult. Change*, July 1973, Part II, *21*, 1–84.

K-C. Yeh, "Capital Formation," in A. Eckstein et al., eds., *Economic Trends in Communist China*, Chicago 1968.

A Model of Public Fiscal Behavior in Developing Countries: Aid, Investment, and Taxation

By PETER S. HELLER*

In most less developed countries (*LDCs*), the public sector's role in the planning and implementation of development projects has been considerable. The rising level of public expenditure has been fueled by capital inflows from public and private sources abroad, and by the mobilization of domestic resources through taxation and local borrowing. Recently, the effectiveness of the government's development efforts have been cast in doubt. Critics argue that foreign capital inflows have resulted in increased public or private consumption rather than increased investment, and have contributed less to growth than was anticipated.[1] Others suggest that the higher tax burden has been squandered on nonproductive forms of public consumption (see Stanley Please (1967, 1972) and L. Krishnamurty).

In this paper, I shall examine these issues by developing a cross-section time-series econometric model of the public sector of eleven African countries (Nigeria, Ghana, Zambia, Kenya, Uganda, Tanzania, Malawi, Liberia, Ethiopia, Tunisia, and Morocco). Nine of these are English-speaking or "Anglophone," and seven were former British colonies; thus one would expect strong structural similarities in their budgetary processes. With the exception of Zambia, the economies of these countries are fairly similar in their structure and level of development. Tunisia and Morocco are unusual among "Francophone" countries in the degree that their fiscal system is independent of that of France,[2] and have been included to test for behavioral differences from the other sample countries. The time period is only postindependence.[3]

The econometric model will focus on the interactions among several categories of public expenditure and of domestic and foreign revenue. The model will further distinguish between alternative *types* of aid (grants vs. loans), and alternative *sources* (bilateral vs. multilateral, private vs. public). The results suggest that aid not only increases investment, but simultaneously facilitates a reduction in the level of domestic taxes and borrowing. However, the magnitude of these effects and the precise response of public consumption to aid

---

* University of Michigan. I am grateful for the helpful comments and criticism of E. Berg, H. Brazer, Jane Chalmers, Saul Hymans, Jan Kmenta, Richard Porter, the managing editor, and an anonymous referee. The Center for Research on Economic Development at the University of Michigan provided secretarial and financial assistance.

[1] Much of this literature has been surveyed recently, see Thomas Weisskopf (1972a,b), Gustav Papanek (1973), Keith Griffin, and John Enos and Mohammed Rahman.

[2] In other French African countries, an important share of local expenditure is financed completely out of the French government's budget.

[3] For certain countries, periods of political or military turmoil are also excluded. The time period used includes: Nigeria (1961–66), Zambia (1961–69), Uganda (1962–69), Malawi (1963–69), Liberia (1964–70), Kenya (1962–71), Ethiopia (1964–70), Tanzania (1963–71), Ghana (1964–70), Morocco (1960–70), Tunisia (1960–71). The sources of data are listed in my 1973 paper. Copies of the data series and discussion of the precise series used for each country can be obtained from the author upon request.

varies according to the type of aid. Grants have a stronger pro-consumption bias, whereas loans are more pro-investment. The model also confirms that only a small proportion of marginal tax increases is allocated to investment, with the bulk used for public consumption.

In Section I a theoretical model of public sector behavior is developed, and the data and econometric procedures used are summarized (with a fuller discussion of the econometric issues in the Appendix). In Section II the results are examined, and Section III summarizes their policy and behavioral implications.

I

One approach to an understanding of the fiscal behavior of the public sector is to assume that it reflects the actions of a set of public decision makers (i.e., a Council of Ministers, etc.).[4] We shall assume that they maximize their utility, taking into account (i) alternative uses of public resources such as expenditures for economic growth, for provision of current social and economic services, and for the maintenance of political order and stability; (ii) the distribution of total output between the private and public sector; (iii) alternative modes of domestic financing such as borrowing and taxation; and (iv) alternative types of external assistance, such as grants and loans.

In any period t, assume that the utility function of the decision makers is

$$(1)\quad U = F[I_g, (Y - T), G_c, G_s, B; A_1, A_2]$$

where I_g = public investment expenditure for development purposes

$(Y-T)$ = disposable income in the private sector (which equals gross domestic product Y less tax revenue T)

G_c = "civil" consumption in the public sector

G_s = "socioeconomic" consumption in the public sector

B = the flow of public borrowing from *domestic* sources

A_1 = total foreign grants to the public sector from all sources

A_2 = total foreign loans to the public sector from all sources

Each variable relates to time period t unless otherwise stated. All variables are in real terms.

The delineation of three expenditure categories, I_g, G_c, and G_s, reflects a functional distinction found in the budgets of many African *LDC*s. The capital budget I_g is the public sector's contribution toward the realization of economic growth targets.[5]

Socioeconomic consumption expenditures G_s include all current noncapital expenditures for socioeconomic ends, including expenditures for the staffing of schools, hospitals, and health centers, for the maintenance of roads and communication networks, for the staffing of agricultural extension or agricultural research projects, etc.[6] Theoretically, one might expect that both G_s and I_g have some impact on the rate of economic growth, but in the eyes of public decision makers, G_s is usually not regarded as investment, but as a form of consumption without developmental impact.[7]

---

[4] In this respect, the model is derivative of models used to explain the fiscal behavior of state and local governments in the United States; see Edward Gramlich and L. Krishnamurty.

[5] I_g includes gross capital formation in the public sector (i.e., buildings and construction, transport equipment, draft animals, etc.) and net loans to other sectors of the economy, with the *exception* of capital formation already included in civil consumption as defined in the text.

[6] G_s is defined to include total government consumption as defined in the United Nations' budgetary statistics, less *current* civil consumption expenditure. Thus it includes current expenditure in such public sectors as education, health, transport, and agriculture. It also includes transfers to local government units since a large proportion of such transfers is used to subsidize this type of expenditure.

Civil consumption G_c includes all other public expenditures, much of this relating to the fundamental need of the state to maintain its political existence. It includes expenditures, capital and current, for government administration, for the servicing of public debt, for diplomatic representation, and for preservation of internal and external security through the police, courts, and armed forces. A smaller fraction is for subsidies and transfers to households and other nongovernmental units.[8] In distinguishing G_c from I_g and G_s, we hypothesize that decisions on G_c are both based on a different set of decision rules and have a stronger claim on public resources.

On the revenue side, increases in the tax burden T become increasingly difficult for the public decision maker to espouse, both because of the increased administrative difficulty of collection, and because of the economic cost and political resistance engendered in countries at low income levels. Of primary importance is the assumption that the choice of tax rate is a policy instrument available to public decision makers, so that T is endogenous.[9]

Borrowing from the local capital market constitutes an alternative means of resource mobilization, although it also yields disutility to the public decision maker. The net increase in the public sector's *long-term domestic* debt B is traditionally seen as fiscally irresponsible if it occurs in more than limited amounts, and unless it is used to finance public sector investment.[10] The latter restriction will be introduced below in our constraint equations.

Finally, capital inflows to the public sector from abroad, whether public or private, bilateral or multilateral, grants or loans, are assumed to be exogenous to the public sector. Although it is possible that these governments would reject aid encumbered by onerous political or economic implications, we shall assume that this is uncommon. We shall also assume that the governments are not in a position to increase significantly the level of capital inflows beyond that actually offered. In the estimations, we shall distinguish between (i) total grants and total loans received from *all* sources (public and private); (ii) net grants and net loans received from public sources; and (iii) bilateral and multilateral aid.[11]

It will be assumed that the utility function in (1) takes the form

$$(2) \quad U = \alpha_0 + \alpha_1(I_g - I_g^*) - \frac{\alpha_2}{2}(I_g - I_g^*)^2$$

---

[7] This can be easily discerned by a perusal of the annual budget speeches of the Ministers of Finance in our sample of African countries. See the author (1973).

[8] Although expenditures for the redistribution of income admittedly arise from a different kind of political motivation, these are not yet a major expenditure in the sample of African countries under analysis. It was felt that in a budgetary decision-making framework, the criteria applied to income redistribution would be most similar to that applied to other G_c expenditure.

[9] Total tax revenue includes direct and indirect property tax revenue (income taxes, customs and excise duties, export taxes, mineral royalties, etc.), and revenue from sales of government social services, license fees, interest, dividends and profits of the government, and miscellaneous revenues.

[10] The net increase in domestic debt was used instead of the stock of outstanding public debt obligations for conceptual and practical reasons. Besides the difficulty of obtaining data on the outstanding stock of debt, the budget messages in most countries indicate a conservative fear of excessive borrowing in any period, which, perhaps irrationally, appears unrelated to the level of existing debt.

[11] The variables in (ii) above measure *official* grants and loans, respectively, received from donor nations and multilateral aid agencies, and are derived from the Organization of Economic Cooperation and Development's (*OECD*) statistics. Grants include both "grant-like" flows (net) and net official grants, reparations, and indemnification payments. Loans are net of amortization and interest. The aid measure in (iii) is a breakdown of the same aid by *source*, i.e., bilateral and multilateral. Finally, each country compiles data on total grants and loans received from all foreign sources, *public* and *private* (this includes official aid as well as supplier's credits), and this is the aid measure in (i).

$$- \alpha_3(T - T^*) - \frac{\alpha_4}{2}(T - T^*)^2$$

$$+ \alpha_5(G_c - G_c^*) - \frac{\alpha_6}{2}(G_c - G_c^*)^2$$

$$+ \alpha_7(G_s - G_s^*) - \frac{\alpha_8}{2}(G_s - G_s^*)^2$$

$$- \alpha_9(B - B^*) - \frac{\alpha_{10}}{2}(B - B^*)^2$$

where $\alpha_i \geq 0$ for all i, and where a starred variable indicates a *target* level for each variable. The functional form chosen ensures diminishing marginal utility for each of the variables I_g, G_c, G_s, B, and T as they rise above a level determined jointly by their target level and by the specific set of α parameters for each variable. It also reflects a compromise between the need for heuristic accuracy and the need for an easily estimable functional form with desirable utility function properties. The absence of any interdependence between the policy variables is its primary deficiency.

The target variables in equation (2) are assumed to be determined by the following relationships:

$$(3) \quad I_g^* = \alpha_{11} Y_{t-1} + \alpha_{12} I_p$$

$$(4) \quad T^* = \alpha_{13} Y_t + \alpha_{14} M_{t-1}$$

$$(5) \quad G_c^* = \alpha_{15} G_{c,t-1}$$

$$(6) \quad G_s^* = \alpha_{16} E + \alpha_{17} Y_t + \alpha_{18}(Y_t - Y_{t-1})$$

$$(7) \quad B^* = 0$$

where M = the value of total real imports
E = primary school enrollments (in units of 10,000 students)
I_p = real private investment expenditure

The target level of investment I_g^* is set in the context of a long-term economic development plan and is influenced by the desired rate of economic growth, the perceived role of the public sector in achieving that growth, the absorption capacity of the public sector, and the relative productivity of public sector investment. Due to the absence of comparable annual investment targets in the country development plans, I_g^* was related to instruments with which it would be highly correlated. In a Harrod-Domar framework, one can posit that for a given target growth rate, I_g^* will be positively related to the level of output in the previous period Y_{t-1}, and inversely related to total private sector investment.[12] However, there may also be positive linkages between I_p and I_g if such investments are technologically complementary.[13]

The civil consumption target G_c^* is assumed to be linearly related to its value in the previous period, reflecting the importance attached to a fundamental continuity of these activities. Several instruments were used to determine G_s^*. Since educational expenditure is a substantial fraction of G_s, primary school enrollments were used to measure the level of this activity. This expenditure is also likely to be held to a fairly constant share of total output Y_t. An accelerator mechanism may also boost this expenditure, and hence $(Y_t - Y_{t-1})$ was also tested as an instrument.[14]

The tax target T^* is determined by the anticipated level of total income and by the value of tax handles such as imports or exports. Since the government sector's expenditure program may substantially affect the level and composition of imports in a given period, imports in the previous period were used as one of the instruments for T^*. Finally, it was assumed that *ex ante*, the borrowing target is equal to zero. This would not, of course, preclude a positive level of borrowing.

---

[12] An accelerator model would suggest that the change in total output $(Y_t - Y_{t-1})$ becomes an instrument.

[13] It was also hypothesized that the level of foreign exchange reserves would determine the practicality of a given level of investment, but this variable proved statistically insignificant.

[14] Since technical assistance aid is principally used to finance the staffing of operational positions within the governments of our sample, it could be considered a substitute for the government's own financing of these positions (although it may entail financial obligations, such as counterpart expenditures). The hypothesized inverse relationship between G_s^* and technical assistance was not borne out, with a completely insignificant effect on G_s. However, this may be due to the poor quality and limited amount of data available on technical assistance.

There are both economic and institutional constraints on the set of feasible public sector decisions. The least restrictive assumption is that all revenue inflows are pooled and allocated among all expenditure categories. Specifically,

$$(8)\quad T + B + A_1 + A_2 = I + G_s + G_c$$

This is institutionally unrealistic. Most African *LDC*s not only reject borrowing for current expenditure, but are encouraged by donors to realize public sector savings through a surplus on the recurrent budget, viz., $(T - G_s - G_c) \geq 0$. This would suggest as an alternative constraint set that

$$(9)\quad I_g = B + (1 - \rho_1)T + (A_1 + A_2)$$

$$(10)\quad G_s + G_c = \rho_1 T$$

where $0 \leq \rho_1 \leq 1$. The level of $(1 - \rho_1)$ reflects the government's belief as to the maximum it can realistically "save" from the recurrent budget, and this enters as a constraint on its decisions. It is not an additional policy variable. Constraint set (9) and (10) also imply that aid cannot be used directly for public consumption. Assuming no relaxation of borrowing effort, aid flows would wholly finance investment.

In the radical literature,[15] it is contended that there is greater substitutability and that aid flows can also, *ex post*, be allocated to consumption. This may be expressed either by relating ρ_1 inversely to the level of A_1 and A_2 in a given period, or by including only $100(1 - \rho_2)$ percent of A_1 and $100(1 - \rho_3)$ percent of A_2 in (9) and the residuals in (10).[16] This suggests the following constraint set:

$$(11)\quad I_g = B + (1 - \rho_1)T + (1 - \rho_2)A_1 + (1 - \rho_3)A_2$$

$$(12)\quad G_s + G_c = \rho_1 T + \rho_2 A_1 + \rho_3 A_2$$

Maximization of U with respect to current policy variables I_g, G_s, G_c, T, and B, given levels of A_1 and A_2 and subject to constraint set (11) and (12), yields the following first-order conditions:

$$(13)\quad \frac{\delta U}{\delta I_g} = \alpha_1 - \alpha_2(I_g - I_g^*) + \lambda_1 = 0$$

$$(14)\quad \frac{\delta U}{\delta G_c} = \alpha_5 - \alpha_6(G_c - G_c^*) + \lambda_2 = 0$$

$$(15)\quad \frac{\delta U}{\delta G_s} = \alpha_7 - \alpha_8(G_s - G_s^*) + \lambda_2 = 0$$

$$(16)\quad \frac{\delta U}{\delta T} = -\alpha_3 - \alpha_4(T - T^*) - \lambda_1(1 - \rho_1) - \lambda_2\rho_1 = 0$$

$$(17)\quad \frac{\delta U}{\delta B} = -\alpha_9 - \alpha_{10}(B - B^*) - \lambda_1 = 0$$

$$(18)\quad \frac{\delta U}{\delta \lambda_1} = I_g - B - (1 - \rho_1)T - (1 - \rho_2)A_1 - (1 - \rho_3)A_2 = 0 \quad \text{and}$$

$$(19)\quad \frac{\delta U}{\delta \lambda_2} = G_s + G_c - \rho_1 T - \rho_2 A_1 - \rho_3 A_2 = 0,$$

where λ_1 and λ_2 are the Lagrangian multipliers associated with constraints (11) and (12). Equation set (13) through (19) can be solved to obtain structural equations for the estimation of the parameters of the utility function and the budgetary constraint relationships:

---

[15] See Griffin and Enos, and Weisskopf (1972a,b).

[16] The only constraint on ρ_2 and ρ_3 would be that total investment exceed the level of capital inflow and required counterpart expenditure. Obviously, an aid recipient cannot invest less than its total aid flow and counterpart commitments without facing a serious credibility problem when seeking additional aid. For our sample countries, this proves to be a nonbinding constraint.

$$(20)\quad \begin{pmatrix} G_s \\ G_c \end{pmatrix} = \begin{pmatrix} \iota \\ -\iota \end{pmatrix}\beta_0 + \begin{pmatrix} -G_c^* \\ G_c^* \end{pmatrix}(1-\beta_1) + \begin{pmatrix} G_s^* \\ -G_s^* \end{pmatrix}\beta_1 + \left\{\begin{pmatrix} T \\ 0 \end{pmatrix} + \begin{pmatrix} -T \\ T \end{pmatrix}\beta_1\right\}\rho_1 + \left\{\begin{pmatrix} A_1 \\ 0 \end{pmatrix} + \begin{pmatrix} -A_1 \\ A_1 \end{pmatrix}\beta_1\right\}\rho_2 + \left\{\begin{pmatrix} -A_2 \\ 0 \end{pmatrix} + \begin{pmatrix} -A_2 \\ A_2 \end{pmatrix}\beta_1\right\}\rho_3 + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix}$$

$$(21)\quad T = \frac{\beta_2}{\beta_3} + \frac{\rho_1\alpha_6}{\beta_3}(G_c^* - G_c) + \frac{\alpha_4}{\beta_3}T^* + \frac{\alpha_{10}}{\beta_3}\{[1-\rho_1][I_g - (1-\rho_2)A_1 - (1-\rho_3)A_2]\} + \epsilon_3$$

$$(22)\quad I_g = \frac{(\alpha_1 - \alpha_9)}{(\alpha_2 + \alpha_{10})} + \frac{\alpha_2}{(\alpha_2 + \alpha_{10})}I_g^* + \frac{\alpha_{10}}{(\alpha_2 + \alpha_{10})}\{(1-\rho_1)T + (1-\rho_2)A_1 + (1-\rho_3)A_2\} + \epsilon_4$$

where

$$\beta_1 = \alpha_8/(\alpha_8 + \alpha_6),\ \beta_0 = (\alpha_7 - \alpha_5)/(\alpha_8 + \alpha_6)$$
$$\beta_2 = \{\alpha_5 - \alpha_3 + \alpha_9(1-\rho_1)\}$$
$$\beta_3 = \{\alpha_4 + \alpha_{10}(1-\rho_1)^2\}$$

Also $[\iota]$ is a $(n \text{ x } 1)$ vector of ones and $[0]$ is a $(n \text{ x } 1)$ vector of zeros. The assumed properties of the error term ϵ_i are described in detail in the Appendix. Equations (3) to (6) indicate the instruments used to estimate I_g^*, T^*, G_c^*, and G_s^*.

Equation (20) reflects the symmetrical relationship that exists between the coefficients of the structural equations for the estimation of G_s and G_c. It was estimated first, in order to obtain estimates of the critical parameters, β_1, ρ_1, ρ_2, ρ_3, and α_{15}. Since (20) is non-linear in these parameters, and since we know that $0 \leq \beta_1 \leq 1$, equation (20) was estimated by iterating over values of β_1 between 0 and 1. The value of β_1 chosen is that yielding the *lowest* sum of squared errors in the estimations. The coefficients of the latter three terms in (20) yield estimates of ρ_1, ρ_2, ρ_3; α_{15} is obtained from the coefficient of the G_c^* term which equals $(1-\beta_1)\alpha_{15}$. On this point see equation (5). These parameters may be used to calculate the transformed independent variables in equations (21) and (22); the coefficients of these transformed variables will then yield estimates of other critical parameters of the system, viz., α_6/β_3, α_{10}/β_3, and $[\alpha_{10}/(\alpha_2+\alpha_{10})]$.

Three econometric problems arose in the estimation of this equation system. First, although we assumed that the structural coefficients are the same across countries, there may be differences in the country specific intercepts. Without careful attention to this issue, recent Monte Carlo studies by Marc Nerlove and others indicate that the uncorrected variation of the error term in a pooled cross-section time-series may cause both bias and inefficiency in an ordinary least squares (*OLS*) estimation of the structural coefficients. Using a method suggested by Nerlove and P. A. V. B. Swamy, a generalized least squares (*GLS*) estimation procedure is used to correct for this.[17] Second, if the uncorrected values of the variables are used, there appears to be heteroskedasticity with a higher variance in the estimated residuals for the larger countries in terms of population or income. By transforming the variables to a per capita basis, this problem is ameliorated. Third, since our model is a simultaneous equation system, this suggests the need for a simultaneous equation estimation procedure to obtain consistent estimators for the structural coefficients. Since our equation system is overidentified, two-stage least squares (*2SLS*) is used for each equation. These econometric issues are dealt with more fully in the Appendix.

---

[17] See Nerlove and Pietro Balestra (1966), T. D. Wallace and Ashig Hussain, and G. S. Maddala.

In the estimations, all the data have been deflated to constant 1966 prices, using each country's own *GDP* deflator.[18] To obtain readily interpretable coefficients, all variables have been converted to dollars using the current year exchange rate. Thus variables initially estimated in local currencies are sensitive to any exchange rate fluctuations. This would distort the nature of the observed statistical relationships if it caused variations in variables which were actually unaffected by a devaluation.[19] However, with the exception of Ghana, there were no significant changes in exchange rates during the period, and there were only four devaluations with respect to the dollar among all sample countries.

Finally, the distinction between the fiscal and calendar year results in varying fiscal periods across sample countries.[20] The convention used was to match all variables calculated on a calendar basis (*GDP*, imports, etc.) to the fiscal year beginning in that calendar year. This complicates the interpretation of regression coefficients since for some countries, the public expenditures are themselves calculated on a calendar basis. The justification for this procedure is pragmatic. Obtaining a moving average of calendar year data would have sacrificed at least one degree of freedom for each country, and the conceptual benefits are uncertain.

II. Econometric Results

The estimation of (20)–(22) yields the structural parameters of the utility function (2) and the constraint equations (11) and (12). Reduced form estimates of the impact of foreign capital inflow on the public sector can then be derived. The results indicate the existence of structural differences in the model parameters of the Francophone and Anglophone countries. They also reveal that inclusion of Zambia fundamentally distorts the values of the parameters. In this section, we shall examine these results more closely. The estimated equations are presented in Tables 1 and 2.

A. *Constraint Equation Parameters*

Estimation of equation (20) yields estimates of ρ_1, ρ_2, and ρ_3 for each of the alternative measures of foreign capital inflow (Table 3). Several points can be made. First, our estimate of ρ_1, the proportion of tax revenues that remain within the recurrent budget, is smaller in the Anglophone country sample than in the pooled sample, the latter including the two Francophone countries. In the pooled sample, no more than 17 percent of recurrent tax revenues flow to the investment budget, as contrasted with 22 to 34 percent for the Anglophone countries.

In both samples, there are marked differences between the share of total (public and private) grants (ρ_2) and loans (ρ_3) allocated to the recurrent budget. Whereas total grants are partially allocated to the consumption budget (with ρ_2 falling to between 0.27 and 0.38), the negative values

---

[18] An alternative deflation of the aid data by an export price deflator for the donor countries was attempted, but this led to only negligible differences in the statistical results.

[19] The appropriateness of this conversion depends on (i) the proportion of government purchases of foreign goods and services, and the source of these goods; (ii) whether aid flows are pegged to some target level of real purchasing power or to an absolute level of local or foreign currency; (iii) the frequency and severity of devaluations with respect to each country's primary trading partner; (iv) the degree of bias caused by use of a *GDP* deflator rather than an expenditure specific deflator; and (v) the degree of currency overvaluation and the policy instruments used to ration foreign currency.

[20] All *OECD* aid data are on a *calendar* year basis, whereas aid data derived from individual country statistics are for the fiscal accounting period used by the country.

TABLE 1—CONSUMPTION EXPENDITURE EQUATION; DEPENDENT VARIABLE = $\begin{bmatrix} G_s \\ G_c \end{bmatrix}$

Samples[f] & Aid Variable	Optimal[a] β_1	Constant = $\frac{\alpha_7-\alpha_5}{\alpha_8+\alpha_5}$	$G_{c,t}-1$ $\left(\text{coeff.} = \alpha_{13}(1-\beta_1)\right)$	Primary School Enrollment (in 10,000)	Y_t	(Y_t-Y_{t-1})	Tax Term[b] (coeff. $=\rho_1$)	A_1 Term[c] (coeff. $=\rho_2$)	A_2 Term[d] (coeff. $=\rho_3$)	R^2 (N)	ν[e] or estimation technique
Pooled Samples											
Total Grants (A_1)	.2	−2.68	0.50	0.14	0.03		0.83	0.38	−0.39	0.93	Dummy
& Loans (A_2)		(−4.12)	(3.88)	(1.01)	(2.22)		(18.04)	(0.89)	(1.88)	(112)	ν=.509
Official Grants (A_1)	.3	−3.04	0.50	0.24	0.03		0.91	−0.15	0.04	0.89	Dummy
& Loans (A_2)		(−3.6)	(3.56)	(1.43)	(2.84)		(12.24)	(−0.40)	(0.10)	(106)	ν=.35
Anglophone Sample											
Total Grants (A_1)	.5	−2.0	0.48	−0.14	0.06	−0.05	0.73	0.27	−0.40	0.92	Dummy
& Loans (A_2)		(−5.4)	(4.70)	(−1.21)	(6.59)	(−2.77)	(21.2)	(1.27)	(−1.64)	(88)	ν=.509
Official Grants (A_1)	.5	−1.69	0.42	−0.02	0.05		0.78	0.65	−0.39	0.92	Dummy
& Loans (A_2)		(−3.65)	(4.3)	(−0.2)	(4.9)		(17.35)	(2.35)	(−1.74)	(80)	ν=.35
Bilateral (A_1)	.4	−2.4	0.40	0.09	0.04		0.69	−0.04	−0.06	0.88	Dummy
Multilateral (A_2)		(−5.53)	(4.15)	(0.69)	(3.56)		(12.5)	(−0.32)	(−0.11)	(92)	ν=.46

Note: The *t*-statistics are in parentheses.

[a] β_1 is that $0 \leq \beta_1 \leq 1$ (in increments of .1) yielding minimum sum of squared errors.

[b] Where variable $= \left\{ \begin{bmatrix} T \\ 0 \end{bmatrix} + \begin{bmatrix} -T \\ T \end{bmatrix}^{\beta_1} \right\}$

[c] Where variable $= \left\{ \begin{bmatrix} A_1 \\ 0 \end{bmatrix} + \begin{bmatrix} -A_1 \\ A_1 \end{bmatrix}^{\beta_1} \right\}$

[d] Where variable $= \left\{ \begin{bmatrix} A_2 \\ 0 \end{bmatrix} + \begin{bmatrix} -A_2 \\ A_2 \end{bmatrix}^{\beta_1} \right\}$

[e] See Appendix for definition of ν.

[f] Samples exclude Zambia.

of ρ_3 indicate that a dollar of total loans actually pulls *nonloan* resources from the recurrent budget. Little difference emerges between the impact of official and private loans on the recurrent and capital budgets. In the Anglophone sample, *both* official and total loans draw 139 percent of their value to the investment budget ($\rho_3 = -.39$).

However, a higher fraction of official grants remain within the recurrent budget (64 vs. 27 percent). The addition of the two Francophone countries drops ρ_2 below zero (in case 2, $\rho_2 = -.15$). Finally, since bilateral and multilateral aid flows have insignificant values for ρ_2 and ρ_3, respectively, the grant and loan components within each must have offsetting influences.

When we included Zambia in the sample, the results were considerably different with more than 100 percent of total grants allocated to the consumption budget, and 172 percent of loans to the investment budget ($\rho_3 = -.72$). Since ρ_2 and ρ_3 are not significantly different from zero for official grants and loans in the samples which include Zambia, it is clear that *private* capital flows account for the strongly divergent effects of the total capital flow measures. The distorting effect of Zambia was also clear from the other equations and parameter estimates. This suggested that the structural dissimilarities in Zambia's industrial structure vis-à-vis the other sample countries has led to a marked difference in its fiscal position and in the character of its fiscal behavior.[21] Although I have not included the cum-Zambian results in the tables, I shall point out the pattern of divergence suggested by the results.

B. *Utility Function Parameters*

Estimation of (20) also yields estimates of the relative values of α_6 and α_8, the quadratic terms in the utility function associated with G_c and G_s, respectively. The lower the value of (α_8/α_6), the greater the disutility arising from sharper deviations from the expenditure target for G_c relative

---

[21] For example, with the exception of 1970, Zambia experienced a net outflow of multilateral assistance throughout the decade, yielding a strong negative multiplicative dummy in the estimation of (20).

TABLE 2

Sample Used and Aid Variable	Constant	Y_t	*var* 1[a] $\left(\text{coeff.}=\frac{\alpha_6}{\beta_2}\right)$	*var* 2[b] $\left(\text{coeff.}=\frac{\alpha_{10}}{\beta_2}\right)$	Imports_{t-1}	R^2 (N)	ν or estimation technique
			Revenue Equations[c]				
Pooled Sample							
Total Grants (A_1)	−3.39	0.24	0.70	1.10	−0.13	0.86	0.51
& Loans (A_2)	(−3.41)	(13.24)	(3.34)	(1.61)	(−2.71)	(56)	
Official Grants (A_1)	−1.74	0.19	0.59	5.21	−0.04	0.88	0.45
& Loans (A_2)	(−2.00)	(11.72)	(3.22)	(3.46)	(−0.89)	(53)	
Anglophone							
Total Grants (A_1)	−4.42	0.26	1.33	1.87	−0.16	0.87	0.51
& Loans (A_2)	(−3.87)	(8.48)	(4.15)	(3.34)	(−2.68)	(44)	
Official Grants (A_1)	−4.83	0.26	1.09	2.44	−0.11	0.89	0.45
& Loans (A_2)	(−4.58)	(9.23)	(3.80)	(3.69)	(−2.20)	(40)	
Bilateral (A_1)	−4.80	0.29	1.58	1.56	−0.10	0.85	0.58
& Multilateral (A_2)	(−4.32)	(9.49)	(3.31)	(3.22)	(−1.61)	(46)	

	Constant	Independent *var* 3[d] $\left(\text{coeff.}=\frac{\alpha_{10}}{\alpha_{10}+\alpha_2}\right)$	Y_{t-1}	R^2 (N)	ν or estimation technique
		Investment Equations[c]			
Pooled Sample					
Total Grants (A_1)	−0.69	0.55	0.03	0.52	0.57
& Loans (A_2)	(−0.78)	(5.08)	(1.99)	(61)	
Official Grants (A_1)	−1.64	0.56	0.05	0.63	0.65
& Loans (A_2)	(−2.15)	(4.83)	(3.72)	(54)	
Anglophone					
Total Grants (A_1)	−0.33	0.48	0.02	0.38	0.57
& Loans (A_2)	(−0.41)	(3.77)	(1.74)	(49)	
Official Grants (A_1)	−0.95	0.31	0.03	0.54	0.65
& Loans (A_2)	(−1.31)	(3.62)	(2.62)	(41)	
Bilateral (A_1)	−0.29	0.38	0.02	0.43	0.57
& Multilateral (A_2)	(−0.42)	(4.63)	(1.96)	(52)	

[a] *var* $1=(\rho_1\alpha_{15}G_{o,t-1}-\rho_1G_o)$

[b] *var* $2=(1-\rho_1)[I_g-(1-\rho_2)A_1-(1-\rho_3)A_2]$

[c] Where values of α_{15}, ρ_1, ρ_2, ρ_3 are derived from coefficients in corresponding equations in Table 1.

[d] *var* $3=(1-\rho_1)T+(1-\rho_2)A_1+(1-\rho_3)A_2$

to that for G_s. In general (α_8/α_6) is equal to 1 for the Anglophone countries and is less than 1 for the pooled sample. Hence our hypothesis that G_o expenditures have higher priority but lower variability relative to G_s is borne out *only* for the Francophone countries. In the Anglophone countries, the rate of diminishing marginal utility is the same for both expenditures. The only difference is that with $\alpha_5>\alpha_7$ (from the negative constant term of these equations), higher utility will be obtained from overachievement of G_c relative to G_s, in terms of their respective targets.[22]

The estimation of equation (21) (see Table 2) yields estimates of the relative values of α_4, α_6, and α_{10}, where α_4 and α_{10}

---

[22] This assumes that both α_5 and α_7 are also nonnegative. This condition cannot be explicitly verified from the results.

TABLE 3—PARAMETERS ASSOCIATED WITH THE UTILITY FUNCTION AND CONSTRAINT EQUATIONS[c,d]

Sample (with aid variable used)	Tax Parameter ρ_1	Aid Parameters ρ_2	ρ_3	Some Utility Parameters $\frac{\alpha_2}{\alpha_{10}}$	$\frac{\alpha_8}{\alpha_4}$	$\frac{\alpha_4}{\alpha_{10}}$	$\frac{\alpha_6}{\alpha_{10}}$	$\frac{\alpha_4}{\alpha_2}$	$\frac{\alpha_6}{\alpha_2}$	Partial Tax Coefficients $\frac{\partial B}{\partial T}$	$\frac{\partial G_s}{\partial T}$	$\frac{\partial G_c}{\partial T}$	$\frac{\partial I}{\partial T}$
Pooled Sample[a]													
Total Grants (ρ_2) and Loans (ρ_3)	.83	.38	−.39	.82	.25	.88	.63	1.08	0.19	−0.08	0.67	0.17	0.09
Official Grants (ρ_2) and Loans (ρ_3)	.91	−.15	.04	.79	.43	.18	.11	.24	0.06	−0.04	0.64	0.27	0.05
Anglophone Sample[b]													
Total Grants (ρ_2) and Loans (ρ_3)	.72	.27	−.39	1.13	1.0	.46	.71	.40	0.63	−0.15	0.36	0.36	0.13
Official Grants (ρ_2) and Loans (ρ_3)	.78	.64	−.39	2.23	1.0	.36	.45	.16	0.20	−0.15	0.39	0.39	0.07
Bilateral (ρ_2) and multilateral (ρ_3)	.66	.02	.07	1.63	1.0	.52	1.00	.32	0.62	−0.21	0.33	0.33	0.13

[a] Excludes Zambia, but includes Nigeria, Uganda, Malawi, Liberia, Kenya, Tanzania, Ghana, Morocco, Tunisia, Ethiopia.
[b] As in (a) but also excludes Morocco and Tunisia.
[c] $U=\alpha_0+\alpha_1(I_g-I_g^*)-\alpha_2/2(I_g-I_g^*)^2-\alpha_3(T-T^*)-\alpha_4/2(T-T^*)^2+\alpha_5(G_c-G_c^*)-\alpha_6/2(G_c-G_c^*)^2+\alpha_7(G_s-G_s^*)-\alpha_8/2(G_s-G_s^*)^2$ $-\alpha_9(B-B^*)-\alpha_{10}/2(B-B^*)^2$
[d] $I_g=B+(1-\rho_1)T+(1-\rho_2)A_1+(1-\rho_3)A_2;\quad G_s+G_c=\rho_1T+\rho_2A_1+\rho_3A_2$

are the quadratic terms associated with tax revenues and borrowing, respectively (see Table 1).[23] For example, since α_6 and α_8 are each less than α_{10}, the results imply that additional borrowing incurs a sharper increase in disutility than the marginal gain in utility from additional units of G_c or G_s. Conversely in the Anglophone cases the additional utility from G_s or G_c falls off more sharply relative to the increase in disutility from taxation (α_6/α_4, $\alpha_8/\alpha_4 \geq 1$).

Since (α_4/α_{10}) is less than 1 in all cases, the marginal disutility of additional taxation increases at a rate less than that associated with increased borrowing. It is interesting to note the impact of including Zambia on these parameters. The ratio (α_4/α_{10}) falls substantially to under .10, indicating that copper-rich Zambia is far more averse to borrowing and amenable to increased taxation than the other African countries. Lacking the tax revenue base provided by abundant mineral resources, the cost associated with increased taxation to the latter countries is, not surprisingly, greater.

Finally, from the coefficients of equation (22) (see Table 2), one can derive an estimate of the ratio of α_2 to α_{10}. For the Francophone countries, (α_2/α_{10}) is less than 1, indicating greater disutility to increased borrowing relative to the diminished marginal utility associated with increased investment. The converse appears to hold for the Anglophone countries.

Given a consistent set of parameter estimates for the $[\rho]$ set and the quadratic terms of (1), one can also obtain the partial derivatives of taxation on borrowing and expenditure. These indicate the impact of an exogenous change in tax revenues on public sector fiscal behavior. In Table 3, it is clear that the bulk of any tax increase is allocated to consumption expenditures, with no more than 13 percent going to investment and 10 to 21 percent going to reductions in domestic borrowing. In the Anglophone sample, the two types of consumption are neutrally affected, with a stronger impact on investment and borrowing. The addition of the French countries weakens the borrowing and investment impact, with a larger share of resources going to consumption, particularly G_s.

These results also verify the fiscal interdependence between the recurrent and capital budgets. If there were complete separability of these budgets, ρ_1 would equal 1 and tax revenues would have no impact on the investment equation and vice versa; both possibilities can be clearly rejected.[24] The opposite extreme of complete fungibility between all revenue sources and expenditures appears equally invalid. Tax revenues and borrowing are not fully substitutable.

---

[23] The coefficients of equation (21) yield estimates of α_6/β_3 and α_{10}/β_3, where $\beta_3=\alpha_4+\alpha_{10}(1-\rho_1)^2$. This allows estimation of (α_6/α_{10}), (α_8/α_4), and (α_4/α_{10}).

C. *Target Parameters*

In addition to the utility and constraint equation parameters, we can obtain estimates of the target coefficients specified in (3) through (7). In the consumption equation (20), the coefficients of $G_{c,t-1}$, primary school enrollment, income, and change in income variables can be used to derive α_{15} through α_{18}, respectively.[25] The lagged $G_{c,t-1}$ term is clearly significant, and yields values of α_{15} equalling .62–.70 in the pooled sample and .83–.96 in the Anglophone sample.[26] Assuming positive values of α_5, these are reasonable values, since utility maximization would yield an optimal value of G_c above $\alpha_{15}G_{c,t-1}$. Income is equally significant with values of α_{17} between .10 and .15, which is close to the average share of G_s in total income. Primary enrollment is positively correlated with socioeconomic consumption, but is not significant at a 10 percent level. Finally, in most cases, the change in income variable is insignificant, and has little impact on any of the other coefficients in the consumption equation. Where significant, it enters with a paradoxically negative sign.[27]

In the revenue equation, income and lagged imports are used to estimate T^*. The income coefficient α_{13} is approximately .20–.30, and is clearly significant. Lagged imports have a surprisingly negative coefficient, which can perhaps be explained by the collinearity of Y_t and Y_{t-1}. Alternatively, high imports in the previous period may relax the need for import tariff revenue reliance in the present. Finally, the investment equations were estimated with and without private investment, since the shortage of data on private investment significantly reduces the sample size.[28] The private investment instrument was clearly insignificant for the Anglophone sample, but quite significant with a positive sign when Tunisia and Morocco are added (with α_{12} in the latter case equalling .24–.26). The other instrument, Y_{t-1}, was significant in all cases with α_{11} equalling 0.02–0.05.

D. *Reduced Form Estimates*

The first-order conditions (13)–(19) can be solved to obtain reduced form equations. Table 4 presents the reduced form estimates of the impact of aid on each fiscal variable. For example, grants (A_1) lead to an increase in total spending by 30 to 60 percent of the grant proceeds. The distribution of this income across fiscal variables markedly differs depending on the type of grant flow and country sample. In the Anglophone sample, *official* grants disproportionately increase consumption, with 9 percent of the grant allocated to investment relative to 40 percent for consumption. The Francophone countries differ substantially with a stronger impact on

---

[24] From the above estimates, $\partial I/\partial T$ ranges from .05−.13 for the non-Zambian samples; $\partial T/\partial I$ ranges from .2 to .5.

[25] For α_{16} through α_{18}, the estimated coefficients need to be divided by β_1; for α_{15}, they must be divided by $(1-\beta_1)$.

[26] From (20), it should be noted that a positive value of α_{15} implies a *positive* correlation with G_c and a *negative* correlation with G_s. Conversely, positive values of α_{16} through α_{18} suggest the reverse.

[27] It was argued that the $(Y_t - Y_{t-1})$ variable would positively affect socioeconomic consumption in the same way as an accelerator might positively stimulate investment. However, from our estimates on aid and investment, it may be that any such positive effect on investment is partly financed by a *reduction* in socioeconomic consumption.

[28] The following countries had data on private foreign investment: Nigeria, Zambia, Malawi, Kenya, Tanzania, Tunisia, Morocco.

TABLE 4—REDUCED FORM ESTIMATES

Sample (with aid variable used)	$\frac{dT}{dA_1}$	$\frac{dB}{dA_1}$	$\frac{dR^a}{dA_1}$	$\frac{dI_g}{dA_1}$	$\frac{dG_s}{dA_1}$	$\frac{dG_c}{dA_1}$	$\frac{dT}{dA_2}$	$\frac{dB}{dA_2}$	$\frac{dR^b}{dA_2}$	$\frac{dI_g}{dA_2}$	$\frac{dG_s}{dA_2}$	$\frac{dG_c}{dA_2}$
Pooled Sample[c]												
Total Grants (A_1) & Loans (A_2)	−.09	−.27	.64	.33	.25	.06	−.07	−.62	.31	.76	−.36	−.09
Official Grants (A_1) & Loans (A_2)	−.19	−.50	.31	.64	−.23	−.10	−.18	−.42	.40	.53	−.09	−.04
Anglophone Sample[c]												
Total Grants (A_1) & Loans (A_2)	−.26	−.35	.39	.31	.04	.04	−.16	−.71	.13	.63	−.25	−.25
Official Grants (A_1) & Loans (A_2)	−.31	−.20	.49	.09	.20	.20	−.27	−.92	−.19	.41	−.30	−.30
Bilateral (A_1) & Multilateral Aid (A_2)	−.26	−.55	.18	.34	−.08	−.08	−.27	−.52	.21	.32	−.05	−.05

[a] $\frac{dR}{dA_1} = 1 + \frac{dT}{dA_1} + \frac{dB}{dA_1} = \frac{dI_g}{dA_1} + \frac{dG_s}{dA_1} + \frac{dG_c}{dA_1}$

[b] $\frac{dR}{dA_2} = 1 + \frac{dT}{dA_2} + \frac{dB}{dA_2} = \frac{dI_g}{dA_2} + \frac{dG_s}{dA_2} + \frac{dG_c}{dA_2}$

[c] Excluding Zambia.

investment. Addition of private grants as reflected in the *total* grant measure had opposite effects in the two samples. In the Anglophone case, it shifts the expenditure increase toward investment; with the Francophone countries included, private grants pull expenditure toward consumption. Thus, both the source of grant and the prevailing budgetary behavior (French system vs. British system) are important determinants of how aid affects the public sector.

As would be expected, the receipt of grants leads to a reduction in taxes and domestic borrowing. The reduction in taxes appears primarily a phenomenon associated with the Anglophone countries. Inclusion of Tunisia and Morocco consistently lowers (dT/dA_1). The decrease in domestic borrowing is less than that of taxes only in the case of official grants received by the Anglophone countries. Inclusion of the two Francophone countries and of private grants strongly leads to a sharper reduction in borrowing than in taxation. Finally, when we include Zambia, the lower ratio of α_4 to α_{10} strongly reverses this pattern, with dT/dA_1 considerably larger than dB/dA_1 (dT/dA_1 equalling $-.30$ to $-.50$ relative to dB/dA_1 close to zero). In other words, when we dropped Zambia, the rise in (α_4/α_{10}) implies a greater unwillingness to either lower or raise taxes relative to the target tax burden. The greater disutility of borrowing is mirrored by the stronger negative effect associated with the receipt of external grants.

Foreign loans contribute to both an increase in total expenditure and a restructuring of the mode of domestic financing and pattern of expenditure. Considerable substitutability exists between domestic and foreign borrowing, dB/dA_2 varying between .42 and .92, with the highest value among the Anglophone countries. In both samples, the impact of loans on taxes is negative but considerably smaller than the impact on borrowing. The result is to increment total spending by less than 40 percent of the total value of the loan assistance.

The loan impact on expenditure is more striking. Loans increase investment with the strongest effect arising from the inclusion of private loans. In both samples, investment rises by 63 to 76 percent of total loans, as contrasted with 41 to 53 percent of official loans. Conversely, both types of consumption expenditure are *adversely* affected by loans. In the Anglophone sample, total consumption falls by 50 to 60 percent of the loan inflow. The Francophone countries reduce their consumption to a lesser degree (and thus reduce taxes and borrowing less as well). The inclusion of Zambia only accentuates these broad trends. The investment impact is larger, as is the corresponding consumption reduction, and there is a sharper decline in domestic tax revenues.

The differential impact of loans and grants is not surprising. Grants have a more stimulative impact on consumption, and a weaker impact on investment. Tax reductions are more likely with an inflow of grants; borrowing reductions more likely in response to a loan inflow. More surprising, *grants* have a more stimulative impact on total spending, which may reflect the unwillingness of African countries to expand public spending through a sharp increase in their debt obligations.

Finally, the reduced form estimates of bilateral and multilateral aid for the Anglophone sample reveal only negligible differences in their respective impact on the public sector. Both cause a small reduction in taxes and a larger reduction in borrowing. Both lead to increases in investment with a small negative impact on consumption.

III. Conclusions

The results obtained from the estimation of the model are instructive as to the structure of the fiscal decision-making process in the public sector of Africa, and they shed light on some recent controversies on aid, taxation, and the public sector. In this section I briefly discuss some policy and behavioral implications of the model.

1) The most reassuring aspect of the results is the confirmation of the most basic model assumptions concerning the sign and magnitude of the utility function and constraint equation parameters. They reveal the strong but not fully substitutable interdependence of the recurrent and capital budgets. Not surprisingly they confirm that public decision makers clearly differentiate in their preferences between types of expenditure and their mode of financing.

In solving for the structural coefficients, the signs are consistent with some basic behavioral assumptions. Increases in tax revenue positively influence the expenditure categories and reduce total borrowing, and vice versa. If the results had shown consistently perverse signs for these relationships, the basic model structure would have been highly suspect. This does not mean that there are no differences in fiscal preference; Zambia and the two French-speaking countries, Tunisia and Morocco, appear to react to fiscal pressures in a different way than the other Anglophone countries.

2) The partial derivatives of taxes on the other expenditure and revenue variables verify Please's contention that an increase in the tax burden is unlikely to be fully used for investment. Far more is likely to be allocated to G_s and G_c. A reversal of the Please Effect would require greater pressure on African governments to change their public sector preference map.

3) The impact of the various aid measures is fully consistent with the assumption of utility maximization, and in particular, with the assumption in (1) that I_g, $(Y-\mathrm{T})$, G_o, and G_s are superior goods, and B an inferior good. The precise effect of aid is clearly determined by the shape of the preference map, and the degree to which the recurrent and capital budgets are interdependent.

4) The attempt to focus on the public sector aspect of the "aid-savings" controversy reveals the complexity of the response mechanism to aid. Foreign loans do *not* fully increment total expenditure, but reduce both borrowing and taxation, which is consistent with the radical critique.

However, aid causes a strong shift away from public consumption and toward investment. Moreover, the reduction in public fiscal effort is primarily focused on reduced borrowing rather than reduced taxation. This is more likely to stimulate the availability of capital for private investment than would the reverse situation. Consequently, if all public domestic borrowing were matched by an equal expansion in private borrowing for investment, the effect of foreign loans on total domestic investment would be considerably higher. However, it may be equally myopic to view this negative effect on consumption as wholly beneficial. If the increase in investment comes at the expense of the maintenance and operating funds for ongoing projects, a loss in the overall marginal productivity of capital may result.

Although grant flows also positively affect investment, there is a much stronger bias toward an increase in public consumption, and indirectly, through tax reduction, of private consumption. The problem in evaluating this grant effect is that even aid proponents will agree that some proportion of aid, and particularly grants, is intended at the outset to be used for government consumption. To support the radical argument, one would have to know the size and sign of the deviation from the unknown proportion. On the other hand, the radical critique is supported by the tax reduction effect.

5) The oft-repeated assertion that a shift from bilateral to multilateral aid will induce a greater increase in investment does not appear to be borne out, at least for the Anglophone countries. Only negligible differences emerge in the pattern of expenditure and revenue changes induced by either type of aid. This does not suggest that a multilateralization of aid will not have a beneficial impact on investment and growth. Rather, it indicates that there are probably considerable differences in the type of aid given by alternative multilateral (or bilateral) agencies.

6) The study suggests that further research on the fiscal activities of the public sector in *LDC*s may be fruitful. Our utility function is relatively simple, being both additive and highly aggregative. A more detailed breakdown of the revenue sources and expenditure uses of the public sector, with a more detailed specification of the determinants of each, would be desirable. By further disaggregation, one could better comprehend the decision processes of the public sector, and the way they are influenced by foreign capital inflow.

APPENDIX

Heteroskedasticity: The structural equations in (20)–(22) were initially estimated using the original values of the variables (deflated and converted to 1966 *U.S.* dollars). In the second stage of the *2SLS* estimations, the estimated variances of the country-specific residuals appeared to be positively correlated with the size of the country in terms of output and population. The problem is corrected by reestimating the equations with the variables expressed on a per capita basis. This clearly lowers the relative dispersion in the variance of the error term for the observations of each country.

Autocorrelation: Testing for autocorrelation is not possible with the usual Durban-Watson statistic, since the data is a pooling of cross-section and time-series data. Theoretically, an estimation of the autocorrelation parameter[29] δ_i for the error term of each of the N countries ($i=1, \ldots, N$) might be appropriate, yielding N parameters to be used in a generalized least square (*GLS*) estimation of the model. For each set of observations, we estimated δ_i and the multiple correlation coefficient of the autocorrelation function for a one-period lag. The correlation coefficient is quite low for most countries ($r<.3$), implying there is probably insufficient autocorrelation to justify a further *GLS* correction of the data. Conceivably our correction for heteroskedasticity reduces the severity of the autocorrelation problem. Since the population series for any country is based on extrapolations of past trends in the birth and mortality rates, the observed growth in population embodies a constant time trend. By dividing by population, one is essentially removing part of any time trend from the data itself.

---

[29] The δ_i is normally referred to as ρ_i in the econometrics literature.

The actual equations estimated are modified in certain cases to remove variables which seemed to induce considerable multicollinearity. Although this undoubtedly introduces specification bias in the coefficients of the included variables, this was judged to be preferable to the indeterminacy which multicollinearity could introduce in the coefficients.

The Pooling of Cross-Section Time-Series Data: The pooling of a cross-section time-series data set necessitates assumptions concerning the stochastic process generating the error term for each equation and as to the similarity of fiscal behavior across countries. We assume the existence of such behavioral similarity, viz., that the structural coefficients of each equation are the same for all countries. Although this is reasonable for the Anglophone countries, the assumption is stretched with the inclusion of Morocco and Tunisia and thus we examine the validity of pooling the entire sample by also estimating separate equations for the Anglophone countries (i.e., excluding Tunisia and Morocco).

If we use an error components approach, the error term for each equation in (20) through (22) is broken down as follows:

$$\epsilon_{it} = u_i + v_{it} \quad \text{for } i = 1, \ldots, N; \quad \text{t} = 1, \ldots, \text{T}$$

where N is the number of countries and T the number of years. We shall also assume that u_i and v_{it} have means of zero, are independent for all i and t, and where

$$E(u_i u_i') = \begin{cases} \sigma_u^2 & \text{for } i = i' \\ 0 & \text{for } i \neq i' \end{cases}$$

$$E(v_{it} v_{it}') = \begin{cases} \sigma_v^2 & \text{for } i = i' \ \text{t} = \text{t}' \\ 0 & \text{otherwise,} \end{cases}$$

$$E(\epsilon\epsilon') = \sigma^2 \begin{bmatrix} A & 0, \ldots, 0 \\ 0 & A \quad\quad 0 \\ 0 & 0, \ldots, A \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & \nu \ldots \nu \\ \nu & 1 \ldots \nu \\ \nu & \nu \ldots 1 \end{bmatrix}$$

where $\sigma^2 = \sigma_u^2 + \sigma_v^2$ and $\nu = \sigma_u^2/\sigma^2$.[30]

A priori one would expect this decomposition of the error ϵ_{it} to be reasonable. Specifically, despite behavioral similarity (viz., equal slopes), there may be differences in each country's intercepts u_i (reflecting different levels of development). However, the choice of estimation procedure will depend on whether u_i is assumed to be random (as above). If one expects that the country-specific intercepts are ***unchanged*** as new points are added to each country's data set, one could assume u_i to be nonrandom and estimable along with the other parameters using a dummy variable procedure. On the other hand, one might expect that u_i is random, and in this case, the Monte Carlo studies of Nerlove, Swamy, and others indicate that using dummy variables may be inefficient in terms of the relative mean square error of the estimators $\hat{\beta}$ (the remaining structural coefficients) and the degree of small-sample bias.[31]

One method suggested to determine the appropriate estimation procedure is to estimate $\hat{\nu}$, according to methods developed by Nerlove and Swamy. Swamy's method (see chapters 2 and 3) requires a specific transformation of the *error sum of squares* obtained in two regressions: (i) on the "group mean" data, and (ii) on the "deviations from the group mean" data. For a given equation ((20), (21), etc.), the corresponding error sum of squares is obtained from this pair of regressions run on the second stage of the *2SLS* estimation.

---

[30] In the econometrics literature, this is referred to as $\rho = \sigma_u^2/(\sigma_u^2 + \sigma_v^2)$, but in order to avoid confusion with our model parameter ρ, we shall use ν.

[31] It should be noted that these Monte Carlo studies focused on this problem in a single equation context. It is not clear whether application to a simultaneous equation problem will change the conclusions significantly. See Maddala.

The Nerlove procedure obtains $\hat{\rho}$ by estimating the variance σ_u^2 of the dummy term coefficients as a proportion of σ^2. The latter equals σ_{ui}^2 *plus* the remaining variance σ_{vi}^2 in the dummy variable equation. Since the Swamy and Nerlove estimates of $\hat{\rho}$ prove considerably different in several cases, $\hat{\rho}$ was calculated as the average of the alternative estimates. Given alternative values for $\hat{\rho}$, the number of time periods, and the number of countries, the Monte Carlo studies of Nerlove and Swamy provide small-sample measures of the relative efficiency of *OLS*, *OLS* with dummy variables, and the *GLS* procedure using $\hat{\rho}$.

In general, if $\hat{\rho}$ is close to 1, the use of *OLS* with dummy variables is the best estimation procedure; if $\hat{\rho}$ is close to zero, *OLS* without dummies dominates. I calculated $\hat{\rho}$ values for each equation in (20) through (22), and transformed the variables in the second stage of the *2SLS* estimation procedure according to the procedure indicated as most efficient in the Swamy-Nerlove studies.[32] These transformed variables were then used as inputs to the estimations of equation system (20)–(22). In other words, the variables in each half of the vector in (20) and the whole vector in (21) and (22) were initially transformed according to the Nerlove-Swamy procedure.

---

[32] The values of ρ for each estimation are specified in Tables 1 and 2. If ρ was close to 1, dummy variables were used for that component of the variable.

REFERENCES

T. Balogh, "Multilateral *vs.* Bilateral Aid," in J. Bhagwati, ed., *Foreign Aid,* New York 1971.

E. M. Gramlich, "State and Local Governments and their Budget Constraint," *Int. Econ. Rev.*, June 1969, *10,* 163–82.

K. B. Griffin and J. L. Enos, "Foreign Assistance: Objectives and Consequences," *Econ. Develop. Cult. Change,* Apr. 1970, *18,* 313–37.

P. S. Heller, "An Econometric Analysis of the Fiscal Behavior of the Public Sector in Developing Countries: Aid, Investment and Taxation," Center Res. Econ. Develop., disc. pap. no. 30, Univ. Michigan, Oct. 1973.

———, "The Dynamics of Project Expenditures and the Planning Process: with Reference to Kenya," unpublished doctoral dissertation, Harvard Univ. 1971.

———, "Public Investment in LDC's with Recurrent Cost Constraint: The Kenyan Case," *Quart. J. Econ.*, May 1974, *88,* 251–77.

J. H. Henderson, "Local Government Expenditures: A Social Welfare Analysis," *Rev. Econ. Statist.*, May 1968, *50,* 156–63.

L. Krishnamurty, "Savings and Taxation in Developing Countries: An Empirical Study," Int. Bank Reconst. Develop., econ. dept. pap. no. 23, Aug. 1968.

G. S. Maddala, "The Use of Variance Components Models in Pooling Cross-Section and Time Series Data," *Econometrica,* Mar. 1971, *39,* 341–58.

R. F. Mikesell and J. E. Zinzer, "The Nature of the Savings Function in Developing Countries: A Survey of the Theoretical and Empirical Literature," *J. Econ. Lit.*, Mar. 1973, *11,* 1–26.

M. Nerlove, "Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross Sections," *Econometrica,* Mar. 1971, *39,* 359–82.

——— and P. Balestra, "Pooling Cross-Section and Time-Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas," *Econometrica,* July 1966, *34,* 585–612.

G. F. Papanek, "Aid, Foreign Private Investment, Savings and Growth in Less Developed Countries," *J. Polit. Econ.*, Jan./Feb. 1973, *81,* 120–30.

———, "The Effect of Aid and Other Resource Transfers on Savings and Growth in Less Developed Countries," *Econ. J.*, Sept. 1972, *82,* 934–50.

S. Please, "Saving Through Taxation—Mirage or Reality," *Finance Develop.*, Mar. 1967, *4,* 1–10.

———, "The Please Effect Revisited," Int. Bank Reconst. Develop., econ. dept. pap. no. 82, July 1972.

M. A. Rahman, "Foreign Capital and Domestic Savings: A Test of Haavelmo's Hypothesis with Cross-Country Data," *Rev. Econ. Statist.*, Feb. 1968, *50*, 137–38.

P. A. V. B. Swamy, *Statistical Inference in Random Coefficient Models*, New York 1972.

T. D. Wallace and A. Hussain, "The Use of Error Component Models in Combining Cross Section with Time Series Data," *Econometrica*, Jan. 1969, *37*, 55–72.

T. Weisskopf, (1972a) "The Impact of Foreign Capital Inflow on Domestic Savings in Underdeveloped Countries," *J. Int. Econ.*, Feb. 1972, *2*, 25–38.

———, (1972b) "An Econometric Test of Alternative Constraints on the Growth of Underdeveloped Countries," *Rev. Econ. Statist.*, Feb. 1972, *54*, 67–78.

D. J. Wolfson and P. J. van de Ven, "Problems of Budget Analysis and Treasury Management in French Speaking Africa," unpublished paper, Int. Monet. Fund.

Demand for International Reserves and the Optimum Mix and Speed of Adjustment Policies

By EMIL-MARIA CLAASSEN*

Do central banks behave rationally with respect to their holdings of international reserves? Casual observation would indicate that countries do not take steps to compensate losses in international reserves until their reserves go down below some critical lower level. On the other hand, the behavior of some notorious surplus countries would indicate a nearly insatiable appetite for foreign reserves, somewhat reminiscent of "Mrs. Machlup's Wardrobe Theory of Monetary Reserves" as Robert Clower and Richard Lipsey, p. 587n, termed it. However, there is no need to resort to dilettantish psychiatry if positive economic analysis can explain this asymmetrical behavior. There is some upper level, below infinity, where central bankers do take action to reduce reserves to a target level even though this upper level may be a greater absolute distance away from the desired amount than is the critical lower level. The purpose of this paper is to show one possible optimization procedure of reserve management that would explain less frequent balance-of-payments adjustment policies above the target level of reserves than below it. Furthermore, this optimizing procedure would also explain abrupt, highly discontinuous, as opposed to smooth, balance-of-payments policy changes.

I. The Optimum Combination of Balance-of-Payments Adjustment Policies

The basic context is that of a country which possibly experiences considerable deficits and surpluses due to random disturbances, but is in "fundamental balance-of-payments equilibrium." The actual balance-of-payments surplus or deficit B is composed of a random component ϵ and a desired component B^*, induced by policy actions:

$$B = \epsilon + B^* \tag{1}$$

where B is the actual change in the level of reserves R; that is, $B = \Delta R$. *Starting from an initial position of equilibrium*, $R = R^*$, any deviation ϵ of actual reserves R from desired reserves R^* may be counteracted, partially or totally, by policy actions:

$$B^* = -\gamma\epsilon \qquad 0 \leq \gamma \leq 1 \tag{2}$$

where γ stands for the policy response coefficient equal to $dB^*/d\epsilon$. If we let X stand for a balance-of-payments policy-induced variable, we then can say

$$\gamma = \frac{dB^*}{dX}\,\frac{dX}{d\epsilon} \tag{3}$$

Thereby we discriminate between the impact of the random disturbance on the policy variable, $dX/d\epsilon = \beta$, and the impact of the policy variable on B^*, $dB^*/dX = \alpha$:

$$\gamma = \alpha\beta \tag{3a}$$

We also abstract from the lagged influence of the random disturbance on policy β as well as the lagged influence of policy on reserves α.

All use of a policy variable X in order to influence the balance of payments will be only temporary because of our assumption of a fundamental balance-of-payments equilibrium. The mean value of

$$dX = \beta\epsilon = \gamma\frac{\epsilon}{\alpha} \tag{4}$$

will be zero. The variance of X

$$V(X) = E(dX^2) = \gamma^2\frac{V(\epsilon)}{\alpha^2} \tag{5}$$

expresses the variability of the policy variable, or in terms of the standard deviation:

---

* Professor of economics, universities of Paris-Dauphine (France) and Giessen (Germany). The paper was written in 1973 during a stay as visiting professor at the University of Pittsburgh.

$$S(X) = \gamma \frac{S(\epsilon)}{\alpha} \tag{6}$$

$S(X)$ represents the return of holding reserves since higher reserves decrease the necessity of adopting internal adjustment policies (lower β, and hence, for a given value of α, lower γ), thus decreasing the variability of X.

Whatever tradeoff central bankers may choose between financing a deficit or disposing of a surplus via reserves, on the one hand, and making the adjustments necessary to avoid a change in reserves, on the other hand, one element in their decision is the choice of the adjustment policy(ies) among the menu of available policy actions. The recent models of the demand for reserves developed by Peter Clark and Michael Kelly are only concerned with expenditure-changing policies. Based on these models, the internal adjustment cost consists exclusively of the variability in the level of income. Hence, in terms of equation (6), the policy-induced variable X is nominal income and the coefficient α stands for the marginal propensity to import; accordingly, the internal adjustment cost varies negatively with the openness of the economy. However, a policy of altering income may be extremely costly—particularly for those economies that are not very open—in comparison, for instance, to an interest rate policy. Thus, a more appropriate response to a temporary deficit may be a temporary rise in the interest rate leading to net capital inflows and to a higher transfer of interest payments to foreigners reducing disposable income. The relevance of the openness of the economy would then depend on the capital account, i.e., on the sensitivity of capital flows to changes in the interest rate. While income changes caused by the expenditure-changing policies are some multiple of the desired change in the balance of payments, changes in disposable income may be only a fraction of the desired change in the balance of payments.

The problem of the choice of adjustment policies among the various types of expenditure-changing and expenditure-switching policies in the government's arsenal of possible measures must be solved on the grounds of cost-benefit analysis. The optimal choice could conceivably be different in a deficit than in a surplus situation. Thus, because of reluctance to create any unemployment, an expenditure-increasing policy could be chosen during surpluses and a tight interest rate policy during deficits. On the other hand, several instruments may be used simultaneously in either situation because of diminishing returns to any policy instrument on the basis of absolutely and relatively increasing internal adjustment cost. Such diminishing returns would induce a combination of policy actions. Even without such diminishing returns, each instrument may damage a different group of individuals within the community so that the government may be keen to employ several instruments in order to disperse the "burden of adjustment" equitably in the community.

An alternative and even stronger argument for the optimum combination of adjustment policies can be made outside of the assumed world of certainty of response of the balance of payments to any adjustment policy. William Brainard has shown that under uncertainty if two (n) instruments are available to hit one target it will generally be optimal to use a combination of both (n) instruments. Under uncertainty, the attainment of B^* is then an expected value. Admittedly, the utilization of one policy instrument could attain the target B^* but only with the risk of undershooting or overshooting. In the presence of two policy instruments, the risk will be reduced. Even if the second policy instrument has the same degree of risk as the first, the principle of economies of diversification applies if we assume imperfect correlation between the impacts of the two instruments. Thus the optimum "portfolio" of balance-of-payments adjustment policies will generally include the utilization of all n available instruments.

II. Discontinuous and Asymmetrical Balance-of-Payments Adjustments

In reality, we observe that governments

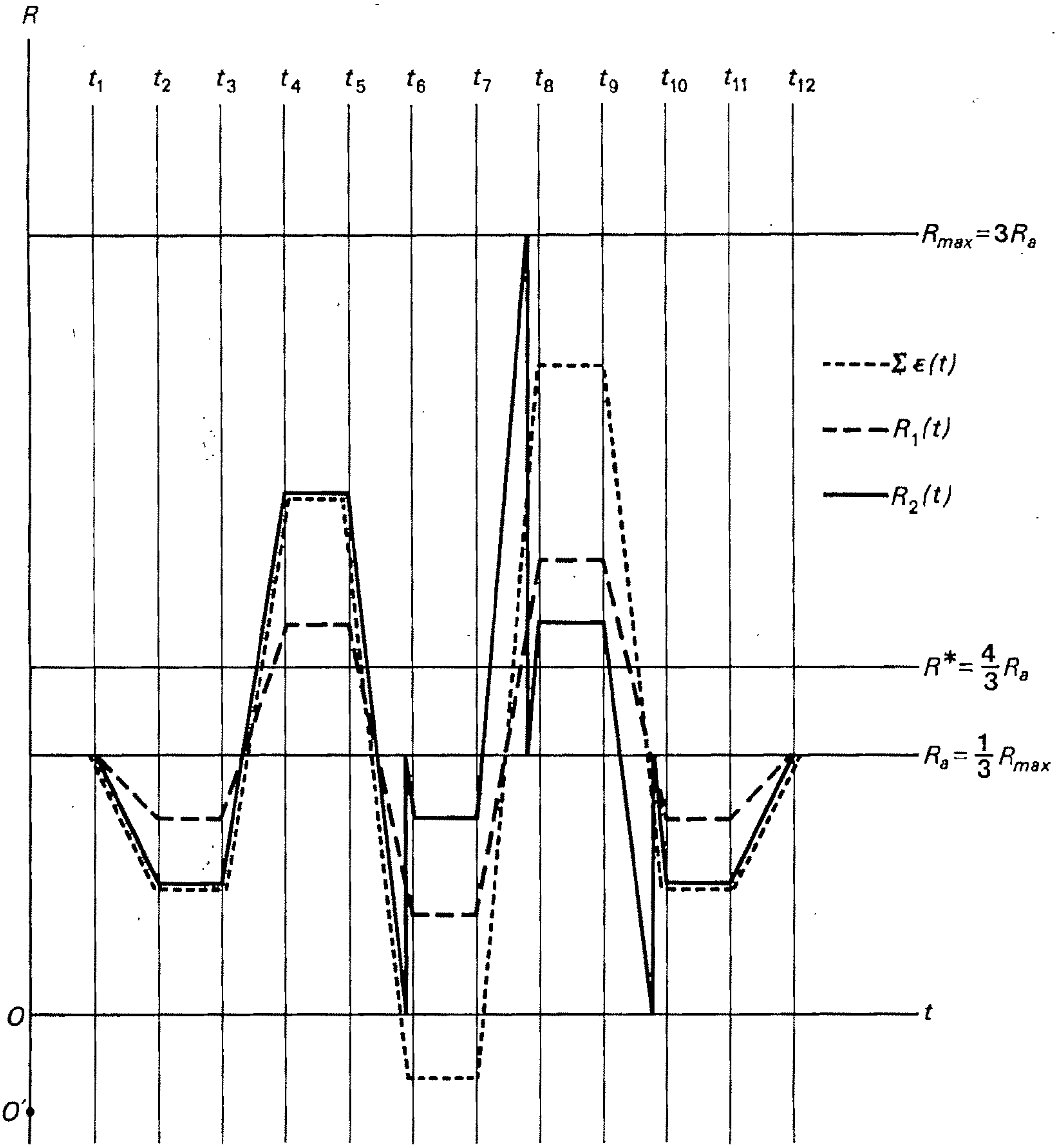


FIGURE 1

not only pursue a combination of balance-of-payments adjustment policies, but that they react discontinuously, and asymmetrically in the cases of surpluses and deficits. Only when reserves go outside certain limits do governments use their policy instruments of adjustment, and they respond faster to a deficit than a surplus.

Figure 1 illustrates one possible speed of adjustment policy in the alternative cases of continuous and discontinuous adjustment. The dotted line indicates one possible time-series of the cumulative random disturbance in the balance of payments, $\sum\epsilon(t)$. Thus, in period t_1t_2 there is a continuous deficit (the exchange rate is at the floor) and the monetary authorities intervene by selling foreign exchange. In period t_2t_3, there is neither a deficit nor a surplus and from t_3 until t_4, the exchange rate is at the ceiling causing renewed intervention by the central bank in the foreign exchange market, this time in order to buy foreign exchange, and so on. If there is a continuous adjustment as the Clark-Kelly model suggests "for analytical sake of simplicity," and this adjustment is a simple linear one (implying that only a fixed proportion of reserves outflows and inflows are neutralized), then the line $R_1(t)$ reflects the behavior of the level of reserves. The vertical difference between $R_1(t)$ and $\sum \epsilon(t)$ measures the continuous correction of deficit or surplus by internal policy actions ($\gamma = 0.5$) on the assumption of the instantaneous influence of these actions on the balance of payments. Line $R_2(t)$ illustrates discontinuous adjustment where the government takes supposedly once-and-for-all internal adjustment measures to bring back the level of reserves to some desired level, say R_a, whenever the reserve level reaches certain lower and upper limits, shown as 0 and R_{max}, respectively, in Figure 1. The assumption that the adjustment policies have an immediate impact on the balance of payments justifies setting the lower signal-point for policy action at zero. A more realistic analysis would set this lower limit at some positive level (say at 0 by shifting the origin of the coordinates to 0′) in the light of problems of uncertainty and possible overshooting and undershooting.

A "discrete-time" adjustment in the sense of less frequent adjustments, but at larger lots, could be justified by pointing out that the random disturbances in one definite direction may soon be reversed so that any "continuous-time" adjustment of the type $R_1(t)$ would be in many cases redundant (and more costly in terms of a higher variability of some internal variables). Only if the disturbance has, for a certain time interval, a definite one-way tendency, will it be necessary to adopt internal policy measures. The main point is that the choice of the "reserves-management rule" (0, R_a, R_{max}) described in Figure 1 can explain the asymmetry in adjustment behavior with respect to surplus and deficit cases, since this rule implies that $R_{max} = 3R_a$.

The above rule can be derived from the analysis of Merton Miller and Daniel Orr of the demand for precautionary cash balances by firms whose payments *and* receipts are timed irregularly (as is the case for central banks). The analysis also involves the assumption that the firms manage their cash balances under some "policy of simple form" (as has been suggested that central banks ought to do). Miller and Orr assume that the probability distribution of the net cash flows has a zero mean, which is in line with our assumption of fundamental balance-of-payments equilibrium. For the hypothesis of a discontinuous cash-management policy (with discontinuous portfolio transfers from bonds to cash or from cash to bonds), they derive two important formulas from probability theory. The first one concerns the probability of a portfolio transfer, $P(A)$—or in our context the probability of a balance-of-payments policy adjustment—and the related question of the probable number of times that reserves (cash) will hit either the upper limit R_{max} or the zero lower limit. The formula for $P(A)$ is:

$$P(A) = \frac{V(\epsilon)}{R_a(R_{max} - R_a)} \tag{7}$$

The second formula concerns the average level of reserves (cash):[1]

$$R = \frac{R_a + R_{max}}{3} \tag{8}$$

There are now two problems to be solved. The first is that of the *relative* level of R_a and R_{max}, or the optimal ratio R_{max}/R_a. The second concerns the *absolute* level of R_a and R_{max}, or the optimum amount of average reserves holdings, R^*. The solution of the first problem is rather easy because it is independent of any return-cost comparison. By definition, R_{max} must exceed R_a; the question is by how much. Suppose that we know the average desired level, R^* (the solution of which will be analyzed later on). According to (8), many possible combinations of R_a, R_{max} will satisfy the equilibrium R^*, and in this respect the ratio R_{max}/R_a is independent on any return-cost comparison given the amount of R^*. Based on (7), the particular combination chosen will set the expression $R_a(R_{max}-R_a)$ at its maximum value, and hence $P(A)$ at its lowest level, subject to the constraint $R^*=(R_a+R_{max})/3$. Condition (8) implies that $R_a(R_{max}-R_a)$ is equal to $3R^*R_a-2R_a^2$; thus we have to maximize this expression with respect to R_a, yielding:

---

[1] The derivation of the above formulas (and possible extensions of them) necessitates a rather lengthy and cumbersome analysis in terms of probability theory. The reader interested in the step-by-step derivation should consult Miller and Orr, pp. 420–23, and Orr, pp. 58–60.

$$R^* = \frac{4}{3} R_a \tag{9}$$

Then setting the value of R^* given by (8) into (9), we obtain:

$$R_{max} = 3R_a \tag{10}$$

The asymmetrical properties of optimal central banks' reserves-management policy then follow. Downward adjustments in reserve holdings, while larger than upward adjustments, are also, of course, less frequent, which is why R^* is not affected by the asymmetrical nature of the solution for R_{max}, given $R_{min}=0$.

We turn to the determination of the average volume of desired reserves R^*, and hence, of the absolute levels of R_a and R_{max}. It can only be solved on the basis of a return-cost comparison. But the return and cost components of reserve holdings have different dimensions. The costs are measured in terms of traditional pecuniary opportunity costs, while the return is expressed by the variability of some internal variables. Therefore one cannot continue the analogy between the optimum amount of reserves and the optimum amount of precautionary cash balances. In contrast to the inventory approach to the demand for money which derives the optimum volume of cash balances by a simple cost-minimization procedure, the demand for reserves is essentially the outcome of the utility-maximization principle.

Let us first look at the "technical" tradeoff between average reserves holdings R and the probable optimal number of adjustments in the reserves $P(A)$. Formula (7) can be rearranged as follows by taking into account our expressions (9) and (10) regarding the relations between R^*, R_a, and R_{max}:

$$P(A) = \frac{V(\epsilon)}{R_a(R_{max}-R_a)} = \frac{V(\epsilon)}{2R_a^2} = \frac{16}{9}\,\frac{V(\epsilon)}{2R^2} \tag{7a}$$

or

$$R = \frac{4}{3}\,\frac{S(\epsilon)}{[2P(A)]^{1/2}} \tag{11}$$

This relationship traces the asymptotic tradeoff curve in the southwest panel of Figure 2. As R increases, the probability of internal adjustment policies to restore reserves to the level R_a becomes smaller.

The optimal point on the tradeoff curve depends on the returns relative to the costs of reserves holdings. The *return* consists of the foregone costs of adjustment, as represented by the variability of some variable X:

$$S(X) = P(A)\frac{S(\epsilon)}{\alpha} \tag{6a}$$

Expression (6a) corresponds to (6) provided that γ is replaced by $P(A)$. However, $P(A)$ continues to be equal to $\alpha\beta$ by analogy with formula (3a). Relation (6a) is represented as a straight line in the southeast panel of Figure 2. Higher reserves lead to less frequent adjustments, thus to the benefit of a lower variability of some internal variables.

The *cost* element of reserves C is equal to the difference between the return the reserves could earn on alternative, domestic or foreign, investments and the positive yield paid on reserves. Denote the net opportunity cost per unit of reserves i; then total cost of reserves will be

$$C = iR \tag{12}$$

The cost function is plotted in the northwest panel of Figure 2.

Now, the technical tradeoff between R and $P(A)$ can be translated into a tradeoff between C and $S(X)$ by setting the value of R given by (11) into (12) and replacing $P(A)$

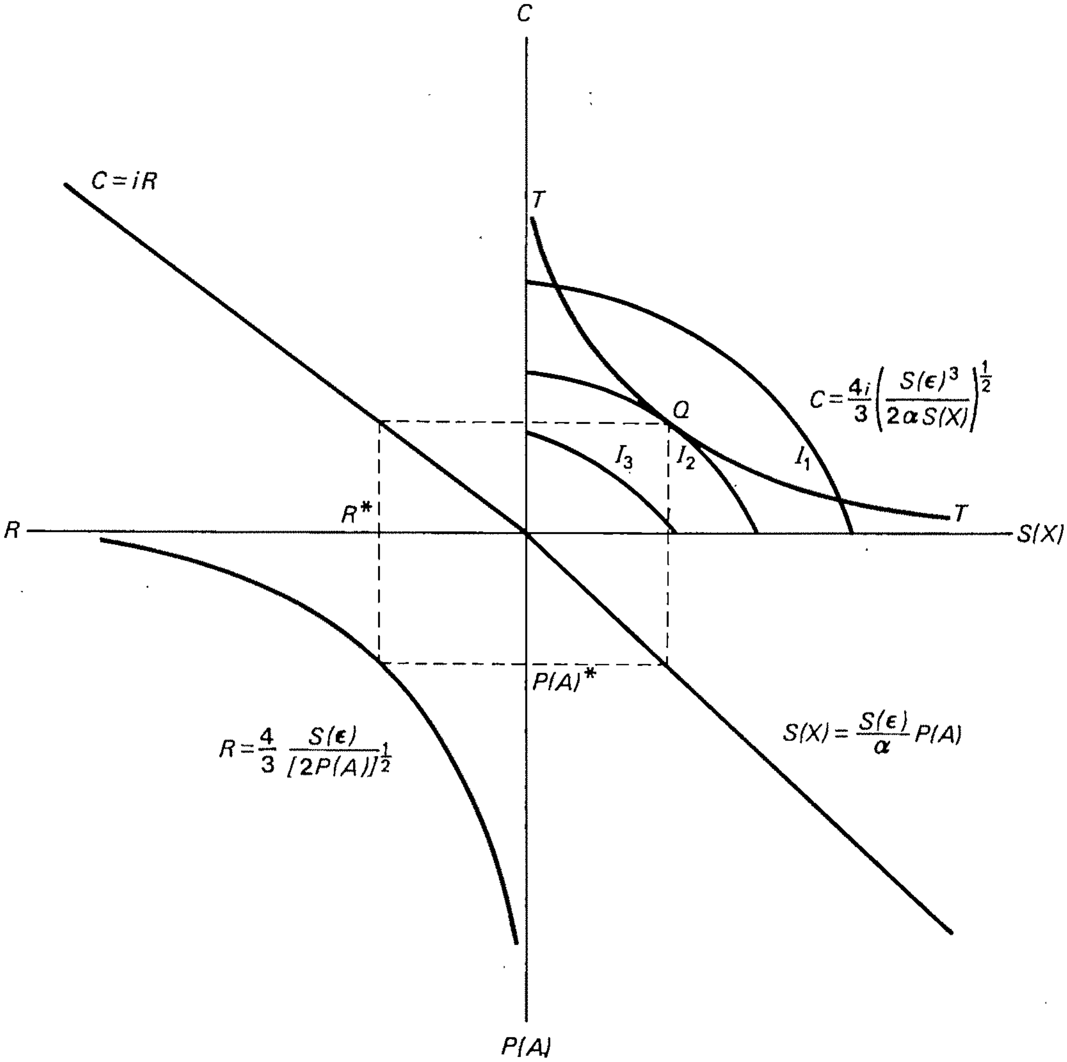

C
C=iR
T
$C=\frac{4i}{3}\left(\frac{S(\epsilon)^3}{2\alpha S(X)}\right)^{\frac{1}{2}}$
Q
I_3
I_2
I_1
T
R
R^*
S(X)
$P(A)^*$
$R=\frac{4}{3}\frac{S(\epsilon)}{[2P(A)]^{\frac{1}{2}}}$
$S(X)=\frac{S(\epsilon)}{\alpha}P(A)$
P(A)

FIGURE 2

by its value indicated in (6a):

$$C=\frac{4i}{3}\frac{S(\epsilon)}{[2P(A)]^{1/2}}=\frac{4i}{3}\left(\frac{S(\epsilon)^3}{2\alpha S(X)}\right)^{1/2} \tag{13}$$

Relation (13) is traced in the northeast panel of Figure 2 as the transformation curve TT. Which cost-return combination the central banker will choose depends on the country's preference function:

$$U=U[C, S(X)] \qquad \frac{\partial U}{\partial C}<0, \frac{\partial U}{\partial S(X)}<0 \tag{14}$$

traced as a set of concave indifference curves in the northeast panel of Figure 2. The shape of the curves assures the attainability of a unique equilibrium at the tangency point Q generating the equilibrium value for the average reserves holdings R^*, and for the rate of balance-of-payments adjustment policies $P(A)^*$.[2]

---

[2] By differentiating (14) and setting $dU=0$ we obtain the slope of an indifference curve I:

$$\frac{dC}{dS(X)}=\frac{\partial U/\partial S(X)}{\partial U/\partial C}$$

This derivation of the "optimum speed" of adjustment policies gives rise to a target level of reserves (R_a), which will be restored discontinuously depending upon when actual reserve movements hit some maximum level (R_{max}) or fall to some minimum level (zero). The maximum level, as shown, is always three times higher than the target level. The *absolute* amount of the target level (R_a), and therefore the absolute amount of the maximum level of reserves and of the average level of reserves (R^*); depends directly on the variability of international payments $S(\epsilon)$ and the internal adjustment cost $S(X)$ and inversely on the interest rate i. The solution also depends on a certain degree of openness of the economy α and a certain "taste" of the "collectivity" regarding the two "bads": domestic income loss due to higher reserves, and internal adjustment cost due to lower reserves.

III. Summary and Conclusion

The suggested model of the demand for reserves is based on government welfare maximizing behavior under uncertainty. Random disturbances in the balance of payments can either be financed with reserves or eliminated by changes of internal variables provoked by so-called balance-of-payments adjustment policies. Thus, there is an "objective" tradeoff between reserves and adjustment policies on the one hand and a "subjective" tradeoff between the opportunity cost of reserves holdings and the variability of internal variables induced by adjustment policies on the other hand. This basic idea can be found in the fundamental contributions of Clark and Kelly. Their basic assumptions, however, are a caricature of actual central bankers' behavior. On the one hand, the monetary authorities are not restricted to expenditure-changing policies as a means of altering reserve holdings. In addition, diminishing returns or increasing disutilities from larger doses of any adjustment policy call for a mixture of available adjustment policies. Uncertainty is another factor breaking down the simple pairing of n instruments with n targets and argues for the efficiency of assigning several adjustment policies to the balance-of-payments target on the basis of the ordinary diversification principle. Another set of issues arises concerning the speed of implementation of this "optimum mix" of adjustment policies, which is neither continuous nor symmetrical above and below the desired level of reserves. It is not irrational for central banks to give high priority to the internal stability target in case of temporarily persistent balance-of-payments surpluses and to give equal priority to internal and external stability targets in case of persistent deficits. Central banks also may be right not only to adjust discontinuously (since adjustment should take place only if actual reserve movements

---

This is the rate at which the country would be willing to accept higher opportunity cost for lower internal adjustment cost. By differentiating (13) with respect to $S(X)$, we obtain the slope of the transformation curve TT:

$$\frac{dC}{dS(X)} = \frac{2i}{3}\left(\frac{S(\epsilon)^3}{2\alpha S(X)^3}\right)^{1/2}$$

which translates the "technical" ratio of the tradeoff between higher opportunity cost and lower adjustment cost. At point Q, the point of the constraint utility maximization, the ratio of the marginal disutilities is equal to the technical ratio of the marginal tradeoff:

$$\frac{\partial U/\partial S(X)}{\partial U/\partial C} = \frac{2i}{3}\left(\frac{S(\epsilon)^3}{2\alpha S(X)^3}\right)^{1/2}$$

A very special case of the country's utility function would be where the marginal rate of substitution of C for $S(X)$ is equal to -1 on each point of the indifference curve. Hence, the indifference curves would be straight lines with a slope of -1. In this special case the equilibrium point would be where:

$$\frac{2i}{3}\left(\frac{S(\epsilon)^3}{2\alpha S(X)^3}\right)^{1/2} = 1$$

This overly restrictive hypothesis would result in a "cube-root law," familiar in the demand for precautionary cash balances and now (partly) applicable to the demand for reserves. In substituting the value of $S(X)$, given by relation (6a) in the above formula, and after having substituted the value of $P(A)$ indicated by equation (7a) in (6a), we obtain the demand-for-reserves function

$$R = S(\epsilon)\left(\frac{16}{9\alpha i}\right)^{1/3}$$

Alternatively, we could have derived the same formula by forming the *total* cost function $K = S(X) + C =$ (6a) $+$ (12) where C is considered as equivalent to the cost dimension $S(X)$. If we differentiate K with respect to R and set $\partial K/\partial R = 0$, we obtain the same cube root formula for the demand for reserves.

hit some ceilings or fall down to some critical minimum), but to respond less frequently, and in larger doses in surplus situations than in deficit situations.

This paper has made extensive application of monetary theory to international monetary phenomena. There is nothing new about such an endeavor even though some authors are not quite conscious about the possible analogies between these two fields of research. There are many similarities between the demand function for precautionary balances by individuals and the demand function for international reserves by central banks. The main difference between the two functions that is recognized in this paper is the use of a notion of internal adjustment cost in the demand for reserves as opposed to the idea of cost of portfolio shifts in the precautionary demand for money. Note, however, that even in modern monetary economics, payments patterns are not always treated as imposed upon the individual. Quite the contrary, in the case of inflation or of interest payments on money, these patterns are considered as highly flexible. The only difficulty in applying the same idea of flexible payment patterns to the central bank demand for reserves is the need to conduct the analysis in terms of social preferences as opposed to simple cost minimization because of the lack of a proper pecuniary dimension for costs.

REFERENCES

W. Brainard, "Uncertainty and the Effectiveness of Policy," *Amer. Econ. Rev. Proc.,* May 1967, *57,* 411–25.

P. B. Clark, "Optimum International Reserves and the Speed of Adjustment," *J. Polit. Econ.,* Mar. 1970, *78,* 356–76.

———, "Interest Payments and the Rate of Return on International Fiat Currency," *Weltwirtsch. Arch.,* no. 4, 1972, *108,* 537–63.

R. Clower and R. Lipsey, "The Present State of International Liquidity Theory," *Amer. Econ. Rev. Proc.,* May 1968, *58,* 586–95.

M. Kelly, "The Demand for International Reserves," *Amer. Econ. Rev.,* Sept. 1970, *60,* 655–67.

M. H. Miller and D. Orr, "A Model of the Demand for Money by Firms," *Quart. J. Econ.,* Aug. 1966, *80,* 413–35.

D. Orr, *Cash Management and the Demand for Money,* New York 1971.

The Hereditary and Relativistic Formulation of the Demand for Money: Circular Reasoning or a Real Structural Relation?

By M. ALLAIS*

The March 1972 issue of this *Review* carries a note by J. L. Scadding on various aspects of the reformulation of the Quantity Theory of Money which I have put forward.[1] Although his comments seem to me to rest on factual errors and paralogisms, they are well worth discussing, for Scadding's view of the basic significance of the new hereditary and relativistic theory seems to be shared by some economists.

It is no easy task to determine the main lines of Scadding's analysis. I think I do him no injustice in singling out the following points. He claims that: a) the weighting system used in the new theory is tantamount to carrying out a lagged regression of the velocity of circulation against itself, which would explain the good results obtained as being a necessary consequence of the formulation, and so destroy any value they might have for analysis of the demand for money. b) It therefore matters little whether one chooses to consider the rate of increase of global expenditure rather than that of prices, since the effect of either tends to "wash out" in the fitting process. c) There is some circularity in the computations, and this would explain the good results obtained. In dealing with Scadding's criticisms, I shall start by recalling the essential features of the new theory (Section I), and go on to review the factual errors in (Section II), and paralogisms (Section III) in his presentation.

I. The Essence of the Hereditary and Relativistic Theory of the Demand for Money

To clarify the critique which follows and avoid possible confusion, a few lines must first be devoted to setting forth the essential contents of the theory I have presented.

1) *The Basic Formulation of the Hereditary Relativistic Theory.* The coefficient of psychological expansion Z is defined as

$$(1)\quad Z(t)=\int_{-\infty}^{t} x(\tau)\exp\left[-\int_{\tau}^{t}\chi(u)du\right]d\tau$$

where $x(t)$ and $\chi(t)$ denote the rate of increase of global expenditure and the rate of forgetfulness at time t, respectively.

The rate of forgetfulness is derived by applying a number of simple and economically appealing economic postulates as

$$(2)\qquad \chi/\chi_0 = 1/\psi(Z) = (1+e^{Z})/2$$

in which χ_0 is the value of χ for $Z=0$. The estimate of M^* (the estimated value of the money supply) is taken as $M^*=\phi_0 D\psi(Z)$ where ϕ_0 is a constant, D is total expenditures.

It should be stressed that the economically significant parameter in this formulation, contrary to Scadding's assumption, is the psychological rate of expansion Z, and not the rate $z=\chi_0 Z$ (Scadding's relation (3), p. 152). Z is an estimate of the cumulative memorized effect of past changes in total expenditure. In no event is it proper to interpret z as "the expected rate of change of . . . nominal output" as Scadding does (p. 151). The only valid interpretation is that it is the rate of change of global expenditure which one is accustomed to. This is a psychological parameter of a hereditary nature, and it has nothing to do with any prediction of the future. More generally, most of the quantities carrying the adjective "expected" in the current literature do not represent in any sense what this label purports to indicate they should.[3] The hereditary and relativistic theory cannot usefully be approached taking currently accepted theories as the frame, for it differs *completely* from them in a number of fundamental respects.

---

* Centre d'Analyse Economique, Paris.

[1] See my 1965, 1966, 1969a,b, 1970, 1972, and 1974 papers.

2) *Fitting Method.* Admittedly, a superficial reading of expression (1) for Z *at first sight* can give the impression of circularity. From (1), Z depends on past values of χ, and from (2), χ is a function of Z. Nevertheless, it is easy to verify that there is no circularity in the fitting process. In fitting, *only* the differential equation

$$(3) \qquad (dZ/dt) + \chi Z = x$$

is used. This is equivalent to relation (1), and the procedure used is to carry out the approximate integration of this relation taking a given value Z_1 as starting point. Thus *for given values of ϕ_0 and Z_1, the computation of M^* depends only on the successive values of global expenditure, and is therefore completely independent of the values of M. The very good fits obtained indicate that there is a complex interdependence between the two series M_n and D_n and that its structure corresponds to that corresponding to the new theory.*

The foregoing shows that appraisal of the scope of the circularity argument derived from considering relation (1) must in fact focus on the differential equation in relation (3), which is the only relation entering the computational process. Furthermore, given the *essential* role played by the values of x in the approximate integration of this equation, it is far from a matter of indifference whether the rate of increase used is that of global expenditure x or prices p.

3) *The Velocity of Circulation.* My own view is that in discussing the hereditary and relativistic theory, it is better to consider the quantities M and M^*, the observed and estimated values of the money supply, rather than V and V^*, the observed and estimated values of the velocity of circulation. Consideration of the series D and V rather than D and M complicates the discussion needlessly, since with $V=D/M$, velocity V depends simultaneously on both D and M. However, as Scadding has brought the discussion onto the latter ground, it is appropriate to consider the implications of the hereditary relativistic theory if the focus of the discussion is V, the velocity of circulation, and not M, the money supply.

To the estimate of M given by $M^*=M_D$, there corresponds the estimate V^* of the velocity of circulation V. We have

$$(4) \qquad V^*/V_0^* = 1/\psi(Z) = (1+e^Z)/2 = \chi/\chi_0$$

where V_0^* represents the value of V for $Z=0$. Equations (3) and (4) show clearly that the calculation of V^* is based not on past values of V but on past values of x. To provide an illustration, Charts I and II present the two fittings for the velocity of money derived using the new theory, for the United States 1918–41 and for Germany from December 1919 to October 1923. Both fittings use exactly the same law, and call on only two arbitrary constants, ϕ_0 and Z_1. The agreement between the calculated and observed values is all the more remarkable for this reason. This agreement should be enough on its own to eliminate any suggestion of spurious correlation.

Putting $v=(V-V_0^*)/V_0^*$ and

$$(5) \qquad v^* = (V^* - V_0^*)/V_0^* = (e^Z - 1)/2$$

we derive from (3) and (4)

$$(6) \qquad [2/(1+2v^*)]dv^*/dt + \chi_0(1+v^*) \log(1+2v^*) = x$$

This equation expresses the complex interdependence linking the values of v^* and those of x. If instead of Z, we consider v^*, the integral equation (1) can be written

$$(7) \qquad \log(1+2v^*) = \int_{-\infty}^{t} x(\tau) \cdot \exp\left[-\int_{\tau}^{t} \chi_0[1+v^*(u)]du\right] d\tau$$

This equation is an integral equation which provides an implicit definition of x as a function of v^*, and vice versa. In terms of the present discussion, (6) is to be preferred to the equivalent relation (7), since it gives the

---

[3] See my 1972 paper, note 22, and 1974 paper, paragraph 10.

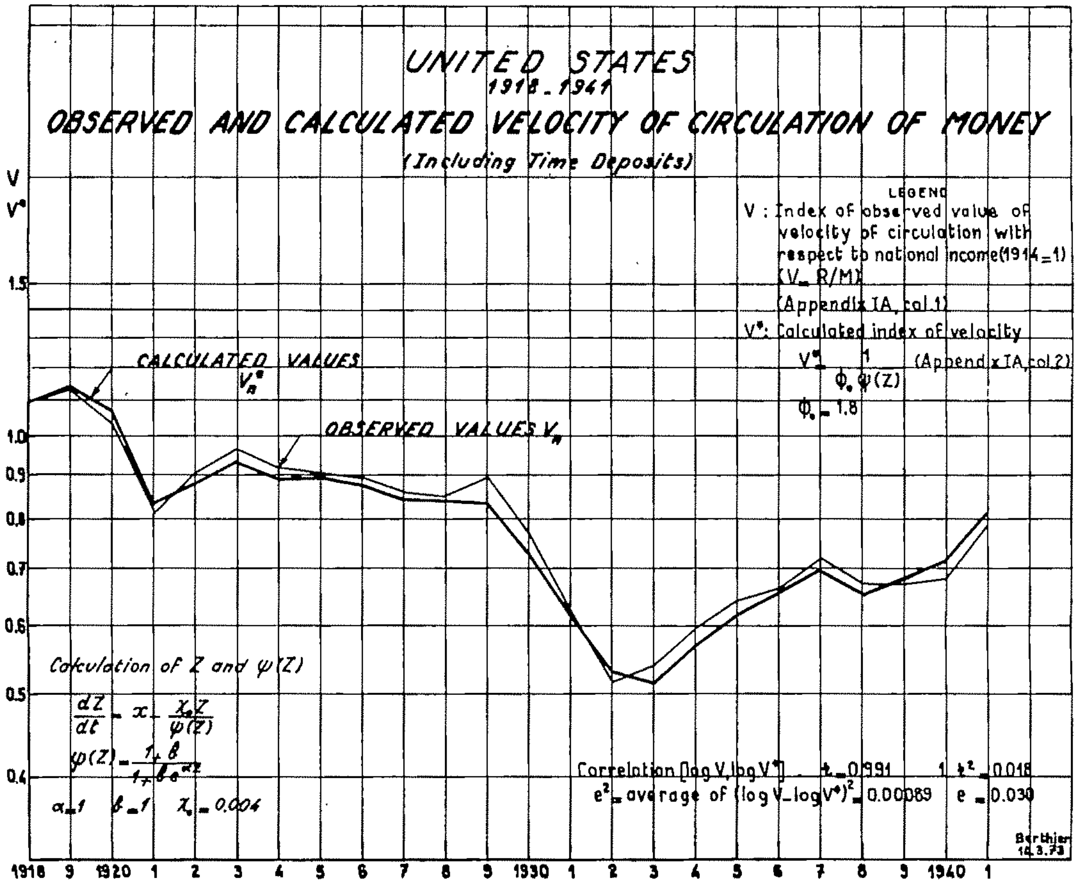


Figure 1

expression of x as an explicit function of v^*, and avoids any appearance of circularity.[3]

II. Estimated and Observed Data—Factual Errors in Scadding's Analysis

The points made by Scadding are based on factual errors which vitiate his reasoning.

1) *Confusion Between the Observed and Estimated Values V and V* of the Velocity of Circulation.* In the first instance, Scadding unfortunately fails to make the distinction between the observed and estimated values V and V^*. In fact, his mathematical reasonings are valid *only if V and v are replaced therein by V* and v**. In using V and v instead of V^* and v^*, he implicitly assumes that the relative deviation of V from V^* remains small[4] which is equivalent to assuming that the hereditary and relativistic theory gives good results. In addition, he makes the further implicit assumption that the same is true for the various higher order derivatives of V and V^*, whereas in point of fact, *the relative deviations are the greater the higher the order of the derivatives considered.*[5] For both of these reasons, Scadding's mathematical

---

[3] At first sight, relation (7) may seem to smack of circularity even more than relation (1). But it must be repeated here, as was noted in Section I.2) above, that the fitting method used demonstrates clearly that there is no circularity (see also Sections III.2) and III.3)).

[4] The implications of this *absolutely essential* assumption are examined in detail in Section III.2).

[5] Since the number N considered by Scadding is very large, there is actually *no justification at all* for treating $D^n v$ and $D^n v^*$ as though they were the same.

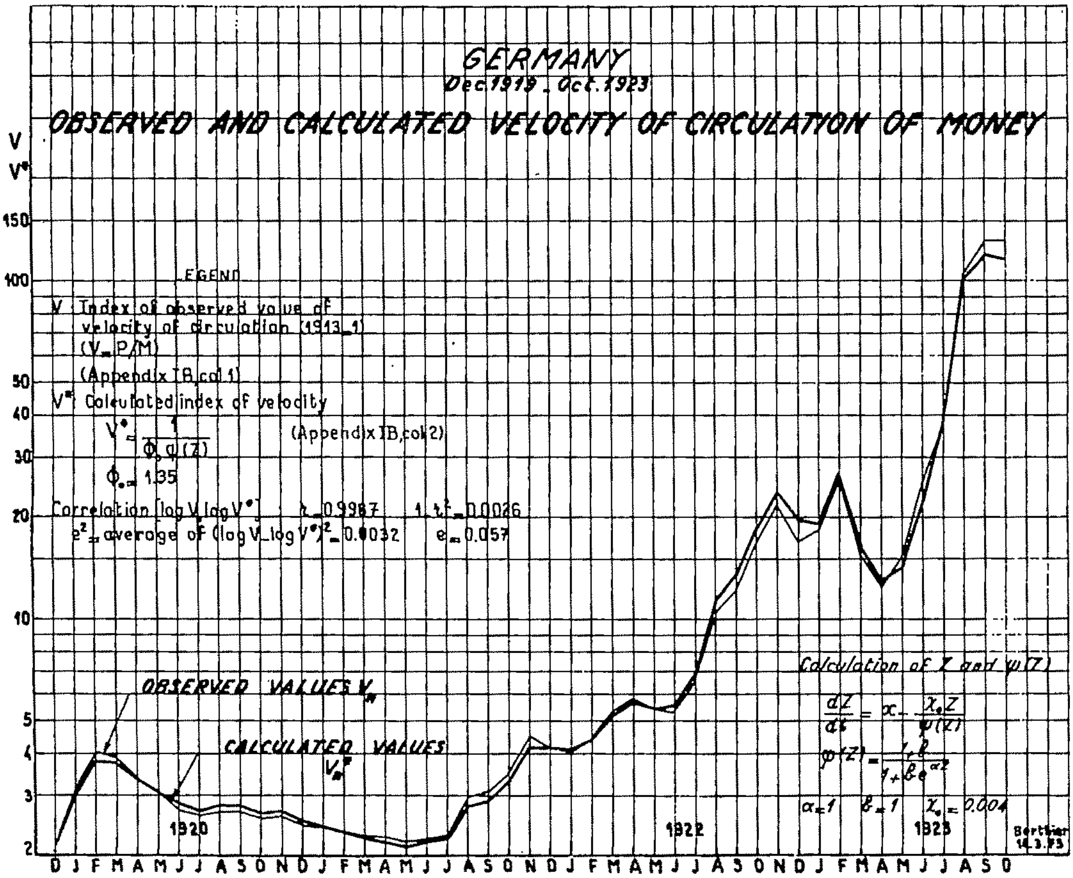


Figure 2

exposition is without foundation.[6]

2) *Unjustified Neglect of Second-Order and Higher Terms.* Secondly, Scadding neglects all second-order terms. As a general rule, such approximation is unacceptable. In fact $v^* = [(1+e^Z)/2]-1 = (Z/2)[1+(Z/2)+(Z^2/6)+\ldots]$ whereas Scadding takes $v^* \sim Z/2$. The relative error affecting v^* by limiting oneself to first-order approximation is shown in Table 1. Thus the relative error ranges from values of 29.7 percent to over 37,150 percent as Z varies from 0.5 to 8.

In fact, Z varies from -0.10 to 1.13 for Chart I, and from 1.5 to 5.78 for Chart II. For Hungary (1945–46) it varies from 3.3 to 8.2. Thus it is clearly wrong to say that "Except in the extremes of hyperinflation the second and higher moments of v should be small enough that we can approximate . . . to a first-order term . . ." (p. 152). This one observation is enough on its own to refute Scadding's reasoning practically in totality, for the validity of estimates, neglecting second-order terms, governs the mathematical basis of his argument.[7] The

---

[6] The confusion between V and V^* (i.e., between M and $M^* = M_D$) is all the more unfortunate in that $M - M_D$ cannot be taken as equal to zero in any way (see my 1972 paper, note 34).

[7] If we write as Scadding does, $\epsilon = (dV^*/V^*)/(dz/z)$ $[z=0]$, and if we consider that $z = \chi_0 Z$ and $V^*/V_0^* = (1+e^Z)/2$, then $\epsilon = (dV^*/V^*)/(dZ/Z) = Ze^Z/(1+e^Z)$ and ϵ is zero for $z=0$. Contrary to Scadding's affirmation, moreover, ϵ is *not* the absolute value of the elasticity of the demand for money with respect to the rate of interest. *This value is unity.* (See my 1969 paper, p. 459, relation (63).)

TABLE 1

Z	0.5	1	2	3	4	5	6	7	8
$\left[\left(\frac{1+e^Z}{2}-1\right)\Big/\frac{Z}{2}\right]-1$	0.297	0.718	2.195	5.362	12.40	28.48	66.07	155.5	371.5

TABLE 2

		United States 1918–1941		Germany Dec. 1919–July 1923	
	Correlations	r	$1-r^2$	r	$1-r^2$
(1, 2)	$\log(V_n/V_{n-1})$, $\log(D_n/D_{n-1})$	0.937	0.123	0.824 (0.485)	0.321 (0.765)
(2, 3)	$\log(V_n^*/V_{n-1}^*)$, $\log(D_n/D_{n-1})$	0.969	0.060	0.837 (0.463)	0.300 (0.785)
(1, 3)	$\log(V_n/V_{n-1})$, $\log(V_n^*/V_{n-1}^*)$	0.966	0.068	0.985 (0.990)	0.031 (0.021)

large changes in Z show how large the variations are in the rate of forgetfulness $\chi=\chi_0(1+e^Z)/2$ and it is impossible, this being so, to consider it as a constant.

3) *Comparative Changes in Global Expenditure and the Velocity of Circulation.* To buttress his arguments, Scadding states: "Moreover the rates of change of nominal output are a relatively smooth series compared to the behavior of velocity" (p. 154). Charts III and IV show that in the case of the series for the United States 1918–41 and Germany 1919–23 used as illustration, *there is nothing of this sort:* the relative changes in x are of roughly the same order of magnitude as those of the derivative $d \log V/dt$.[8] Similar relationships are observed for the other series, whether they relate to current situations or hyperinflations. Correlations of this kind are very remarkable indeed. They must be explained. The hereditary relativistic theory does this perfectly. As a matter of fact, the same correlations are observed to hold more closely between the x and the estimated quantities $d \log V^*/dt$. The data shown in Table 2 are obtained for the United States 1918–41 and Germany, December 1919–July 1923. The figures in brackets for Germany relate to the whole period December 1919–October 1923.

Relation (6) shows that according to the hereditary and relativistic model, the rate of change x of global expenditure, is the sum of two terms, of which the first is proportional to $d \log V^*/dt$, the coefficient of proportionality declining from 2 to 1 as V^* rises from V_0^* to $+\infty$. *This fact makes it possible to explain the correlation observed between x and $d \log V/dt$, insofar as the difference between V and V^* remains small, and insofar as the second term does not take on exceptionally high values*, which it does in the case of Germany during the three final months, from August to October 1923. Because of this, the correlation coefficients for Germany for the whole period December 1919–December 1923 have been shown in brackets: for the last three months, the influence of the second term of the left-hand side of relation (6) is anything but negligible.

Using the subscripts 1, 2, and 3 to denote $\log(V_n/V_{n-1})$, $\log(D_n/D_{n-1})$, and $\log(V_n^*/V_{n-1}^*)$, respectively, the standard formula

$$r_{12.3} = (r_{12} - r_{13}r_{23})/\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}$$

yields values of $r_{12.3}$ for the United States of 0.0076 and for Germany of -0.0011 when applied to the coefficients of Table 2. This suggests the conclusion that the series behave *as if* the observed correlation between

---

[8] In Chart III, x and $d \log V/dt$ are plotted against identical scales. In Chart IV, the scale for x is double that for $d \log V/dt$.

log (V_n/V_{n-1}) and *log* (D_n/D_{n-1}) is entirely due to the correlations between *log* (V^*_n/V^*_{n-1}) and *log* (D_n/D_{n-1}) on the one hand, and *log* (V_n/V_{n-1}) and *log* (V^*_n/V^*_{n-1}), on the other. Since from relation (4)

$$(8) \qquad dZ/dt = (1 + e^{-Z}) d \log V^*/dt$$

the correlation between x and *log* (V^*_n/V^*_{n-1}) is mainly due to the strong correlation between x and dZ/dt, resulting from relation (3) and the very strong correlation between dZ/dt and $d \log V^*/dt$ resulting from relation (8). This correlation evidently becomes stronger still when Z takes on high values, as it does in hyperinflations. It follows that it is essential to take due account of the order of magnitude of the three quantities entering into relation (3).

From an economic standpoint, these results indicate that the hereditary relativistic theory does explain why fluctuations in global expenditure are *always* accompanied by changes in the same direction and of the same order of magnitude in the velocity of circulation. It can also be seen that, contrary to what Scadding seems to suggest, cyclical fluctuations are also to be observed in the German hyperinflation. Their period is much shorter, and their amplitude increases as the hyperinflation proceeds. This phenomenon is also to be observed for the other hyperinflations. It follows that the arguments put forward based on consideration of cyclical fluctuations *should equally apply* to the fluctuations observed during hyperinflations.

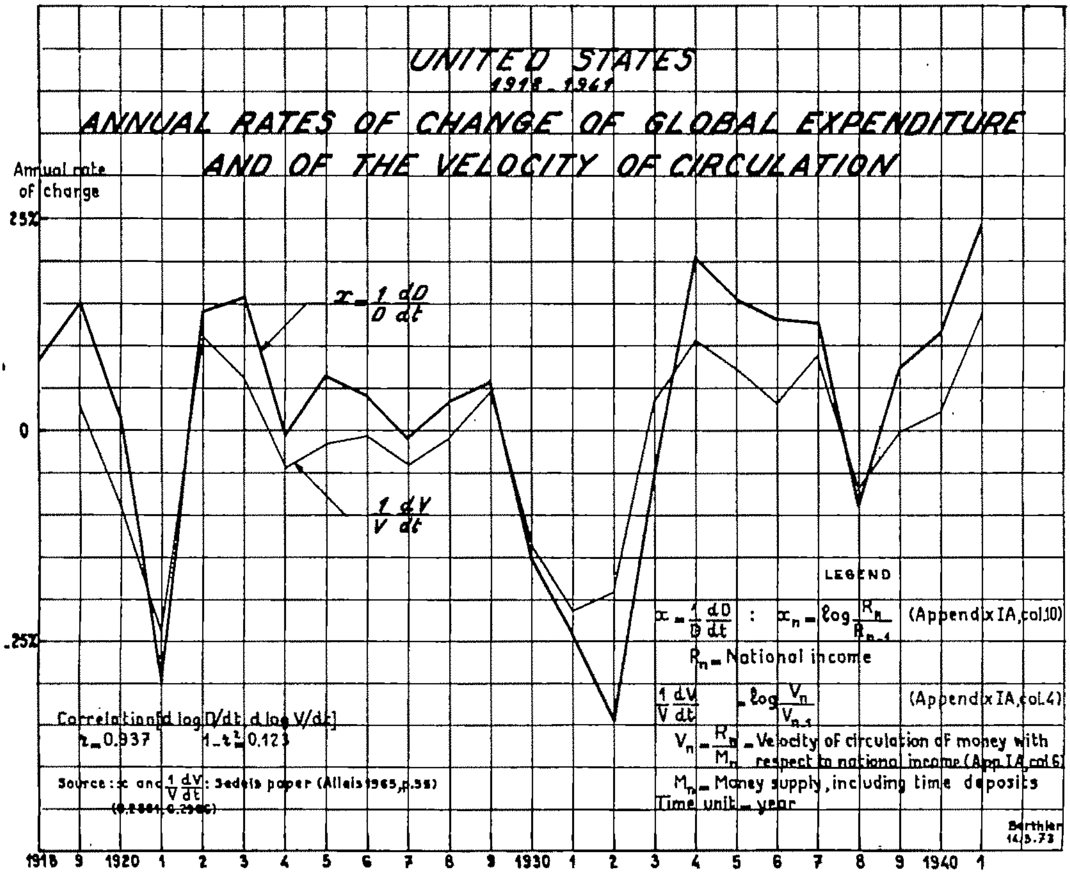


Figure 3

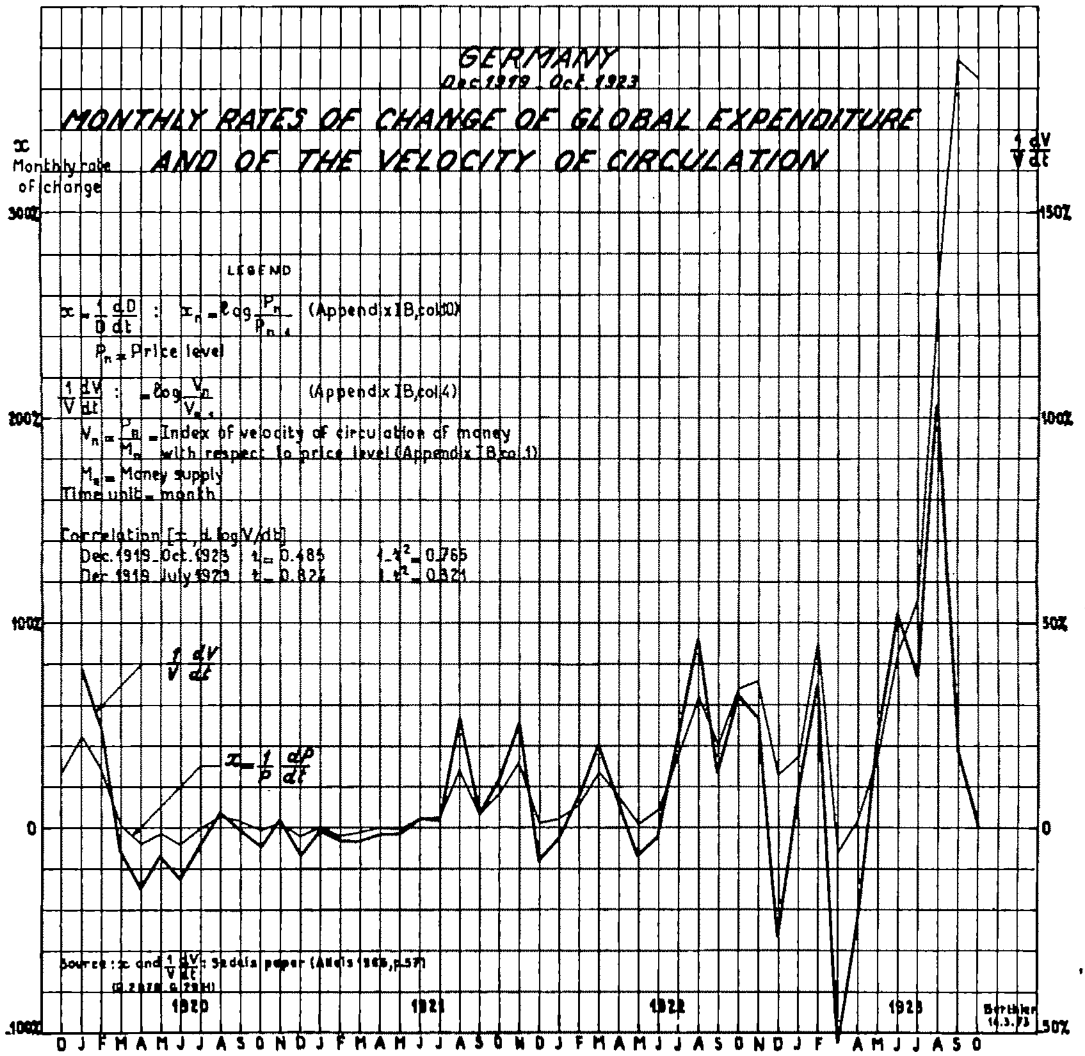


Figure 4

III. Circular Reasoning or Real Structural Relationship—Scadding's Paralogisms

In fact, Scadding's reasoning is founded on a series of paralogisms.[9]

1) *The Term in* z_1. Consider first Scadding's equation (19)

$$
(9)\quad v(t) = b_0 + b_1 z_1(t) - b_1\chi' \int_0^\infty x(t-s) \cdot \left\{ \sum_{n=1}^{N} D^{n-1}v(t-s)\frac{s^n}{n!}\exp\left[-\chi' s\right] \right\} ds
$$

He concludes from this that ". . . the Allais method of estimating velocity comes down to estimating it as a distributed lag in itself and its derivatives" (p. 153).[10,11] Even if this assertion is considered valid as regards the third term on the right-hand side of (9), *which it is not*,[12] it certainly cannot hold for the term in z_1 whose expression

---

[9] So confusedly interwoven are the factual errors and paralogisms in Scadding's paper that it is not easy to separate them out clearly.

[10] Equation (9) is written using Scadding's notation. In point of fact, the term b_0 should be considered as equal to zero (see fn. 11 below). Equation (9) is merely equation (7) above, ignoring second-order and higher terms in v^*, integrating $\int v^*(u)\,du$ by parts and replacing v^* by v. In other words, *if* v is entered in place of v^* in

$$(10)\quad z_1(t) = \chi' \int_{-\infty}^{t} x(\tau) \exp\left[-\chi'(t-\tau)\right] d\tau$$

is given by Scadding's relation (4).[13] Thus there is an error of reasoning here.[14]

No conclusion of any kind can be derived from the expression (9) in v if only the third term on the right-hand side is considered, and the second neglected. It can of course be argued that with x correlated with $d \log V/dt$, the quantity z_1 is itself a weighted average of past rates of change of the velocity of circulation. ***But in this sense, every argument brought forward against the hereditary and relativistic theory also holds against Cagan's theory.*** Be this as it may, the correlation between x and $d \log V/dt$ needs to be explained. It cannot be presupposed as a point of departure.[15] In fact, the "explanation" can be derived only from a model resting on certain assumptions. Why it comes about that such an explanation is possible, i.e., that such a model can actually be developed and verified, is of course a question that lies beyond the domain of science.

2) *The Arbitrary Identification of V and V* in the Reasoning, and its Implications.* The essential paralogism committed by Scadding rests on his arbitrary identification of V with V^* already noted (Section II.1) above). Its meaning and implications must be stressed as follows. Even if the approximations underlying equation (9) are valid, ***which they are not***,[16] equation (9) only applies for the quantity v^*, not for v. For Scadding's analysis to be correct, it would have to be true that $V \equiv V^*$. But this would be equivalent to supposing that the hereditary and relativistic theory represents the reality not only very well, but also perfectly.

In point of fact, Scadding's grounds for the charge of circularity derive only from an absolutely essential assumption which he does not state explicitly. This is that the hereditary and relativistic theory is valid. Obviously under this assumption the conclusion that there is circularity is a necessary corollary, since relation (9) is actually an identity once the relation $v \equiv v^*$ is accepted and second-order terms neglected. But this assumption cannot be accepted without ruining the foundation of all of Scadding's reasoning. Since the equality $V = V^*$, even if only approximate, is tantamount to assuming that two constants Z_1 and $V_0^* = 1/\phi_0$ can be found such that when (3) is integrated using observed values of x and an initial value Z_1 of Z, the values of V^* calculated using (4) will be close to the observed values of V.

In fact, whereas equation (9) with v^* in the first term in place of v follows ***necessarily*** (neglecting second-order terms) from the hereditary and relativistic formulation, there is no necessity at all for constants Z_1 and

---

equation (9) and *if* second-order and higher terms are neglected, this equation is an identity in v^*. In fact, Scadding's relations (10) and (11) are incorrect. The coefficient of the second term on the right is χ'^2 and not χ'.

[11] Scadding's relations (17) and (18) are incorrect. They should be read

$$v = (V - V_0)/V_0 = \{[1 + b \exp(\alpha z/\chi')]/(1+b)\} - 1$$

and, to a first-order approximation $v^* = b\alpha z/\chi'(1+b)$. It follows that, in Scadding's notation, in his equation (19), b_0 should be taken as zero.

[12] See Section III.2) below.

[13] z_1 is the value taken on by the product χZ when χ has the constant value χ_0, the coefficient Z being defined by relation (1) above.

[14] In any case, I do not see what value there is in reducing the hereditary and relativistic formulation to the Allais and Cagan formulation of 1953–54, which is now hopelessly out of date and irrelevant from any standpoint. In fact the real issue is not to try to link up the new formulation, which rests on a variable rate of forgetfulness, with the old one, which rests on a constant rate of forgetfulness. This link follows immediately and has no real interest. The real issue is to choose between the two formulations from an examination of the facts, and not from a priori reasoning.

[15] See Section II.3) above and Sections III.2) and III.3) below.

[16] It should further be recalled that equation (9) is wrong even for v^*, since it neglects second-order terms that cannot be neglected (see Section II.2) above). In reality, it is equation (6) above that should be used, not equation (9). I leave aside the obvious difficulties which arise from the consideration of higher order derivatives of v^*. From equation (6) above, this implies in particular the assumption that the various successive derivatives of $x(t)$ can effectively be considered.

ϕ_0 to exist such that there would be little difference between V and V^*. Alternatively, assuming that second-order terms can be ignored (which they cannot be), equation (9) (written with v^* in place of v) is an integral equation which defines v^* as an implicit function of the x. There is no a priori reason for a solution to be available in which the quantities v and v^* stay close to each other.

The value of the new theory and its empirical application lies precisely in the fact that it shows that such a solution exists, but this does not alter the fact that, a priori, there is no reason why it should. In fact, if $V \equiv V^*$ is assumed a priori, the result is mathematical circularity, whereas the purpose of the analysis is and must be to explain why V and V^* should be approximately equal.

To summarize, from the interdependence between x and V^* specified in relation (6), we cannot derive the conclusion that there is an equivalent interdependence between x and V, except if there is a close correspondence between V and V^*. This is an assumption that cannot be made a priori—without presupposing precisely that the hereditary and relativistic theory is valid. In reality, the implicit confusion of V with V^* has led Scadding to put forward a paralogism, and unfortunately, there is a some presumption that the reader would not perceive this unless forewarned.

3) *Absence of Real Justification for Scadding's Arguments.* Scadding's equation (19), given as equation (9) above, is derived from equation (7) above by replacing v^* by v, neglecting second-order terms and integrating $\int v du$ by parts. If this equation (7) is considered, there is no longer identity. As I have just shown, there is no a priori reason for the values of x estimated from (6) using the values of v instead of v^*, to be identical to the observed values.

To say that equation (9) above consists of regressing the index v on itself is not enough to explain the excellent results provided by the hereditary and relativistic theory, or for that matter, to prove anything at all. This is an affirmation where demonstration is called for. Scadding ought to show that the nature of the regression is such that the second term of equation (9) approximately reproduces the first. Examination of the results of contemporary research shows that lagged regressions do not necessarily always produce good results. When they do, there must be a particular structure which causes them, and it needs to be explained. This is exactly what the hereditary relativistic theory does in the present case.

Autoregression in itself does not guarantee a good fit. Scadding's thesis would mean that any variable whatever could be substituted for x as long as it is correlated with v and that any other function could be used instead of $\psi(Z)$ without any incidence on the good results obtained! For Scadding's reasoning to hold, this would require *either* a mathematical demonstration that the right-hand side of (9) does indeed reproduce the first, at least approximately, for any set of values of x, as long as they are correlated with v; *or* an illustration of his thesis by at least one specific example in which determinate functions are considered for v and x; or even a specific numerical example showing in a particular case that the reasoning presented is justified. The discussion in Section III.2) above shows clearly that this task is impossible. Otherwise, there would be a demonstration that the two sides of relation (6) where v^* is replaced by v necessarily differ little from one another. This is of course impossible unless one has postulated beforehand the complex and remarkable interdependence between the x and the V which the hereditary relativistic theory brings into the open. This postulate is implicit in Scadding's analysis. It follows that Scadding's thesis rests on circular reasoning, and that, for want of support, his thesis breaks down.

In any event, Scadding's arguments apply equally to Cagan's formulation (see Section III.1) above) once the correlation between x and $d \log V/dt$ is assumed a priori. But then what should be explained, and what Scadding does not explain at all, is the reason why one of two theories subjected to the same criticism (groundless in any case) produces markedly better results than the other. In other words, we should be told why consideration of the quantity Z, corresponding to relation (1), gives better results than consideration of z_1, corresponding to relation (10). In my view, the reason is that the hereditary formulation is closer to reality because it considers relative desired money balances M_D/D, instead of real desired money balances M_D/P; the rate of increase of global expenditure x instead of p, the rate of increase of prices; a logistic instead of an exponential function for M_D/D, and a variable rather than a constant rate of forgetfulness. Last but not least, relative desired money balances are assumed invariant when measured against a psychological time scale. It follows that the currently accepted models of the demand for money (whether based on Philip Cagan's rate of anticipation, Milton Friedman's permanent income, or the rate of interest) are all misspecified.

4) *The Correlation between the Rate of Change of Global Expenditure and of the Velocity of Circulation.* Scadding writes: "Since the short-run behavior of velocity is pro-cyclical, the rates of change of nominal output and velocity are positively correlated. Hence the configuration of x-weights will tend to preserve the past variation in velocity," and, a little earlier, "In other words, the sum of terms representing the difference between the Allais and Cagan estimates is a series expansion in v, not x" (p. 154).[17] This leads him to conclude that "For all these reasons, the Allais procedure is approximately that of regressing velocity on a smoothed version of itself" (p. 154).

In the first place, a correlation between x and $d\ log\ V/dt$ does exist, as can be seen from Charts III and IV above for the United States, 1918–41, and Germany, December 1919–October 1923. The correlation coefficients are, respectively, 0.936 ($1-r^2=0.123$) and 0.485 ($1-r^2=0.765$), whereas the correlation coefficients between $log\ V$ and $log\ V^*$ are 0.991 ($1-r^2=0.018$) and 0.998 ($1-r^2=0.0039$), as shown in Charts I and II above. But this makes it somewhat difficult to explain the correlation observed between $log\ V$ and $log\ V^*$ (Charts I and II) as being due to the correlation between x and $d\ log\ V/dt$ (Charts III and IV). In fact, the opposite is true. The correlation between x and $d\ log\ V/dt$ is in reality no more than the consequence in a particular case of the structure of the model represented by equation (6).

Actually, of course, there is no reason for this to be so a priori. As I noted earlier, it is remarkable that the correlation between x and $d\ log\ V/dt$ should not only be found for current situations (what Scadding calls "cyclical"), but also for all the hyperinflations studied. This results from equation (6) above, which shows that the first element of x is proportional to $d\ log\ V/dt$. Thus the hereditary and relativistic theory provides an explanation of this very remarkable fact.

Secondly, it is a paralogism to conclude from the cyclical structure of the time-series that the hereditary and relativistic formulation is tantamount to regressing velocity on itself. The whole point is precisely to determine explicitly the conditions that make this so, and in fact these conditions correspond to a specific structure of the economy which the new theory brings into the open.

Thirdly, Scadding's reasoning *completely* ignores the essential fact that the hereditary and relativistic theory applies equally to the full span of hyperinflationary periods and to current situations. Any attempt to question the economic significance of the new theory should therefore be applicable without modification to hyperinflations. Actually, the remarkably good results yielded by the hereditary and relativistic theory when applied both to hyperinflations and current situations is a fundamental aspect of it which the critics have neglected in a manner which is quite difficult to explain. This in turn suggests that the models used hitherto are misspecified since they are incapable of explaining the complex interdependence of the money supply and global expenditure time-series which is put in evidence by the hereditary and relativistic theory, while at least some of the results (which in any case are not very good) of the currently used models are probably the fruit of spurious correlations.

---

[17] For Scadding's reasoning to hold, this proposition should also apply to Cagan's estimate z_1 (see Section III.1) above).

Scadding also writes: "The fact that extrapolation of time-series often give good predictions may be enough to explain the good results that Allais obtains" (p. 151). It is hard to see how the—not particularly regular—success of extrapolations on short periods can provide any explanation of the success with which fittings were obtained covering normal situations lasting over long periods of years as well as the full duration of a series of hyperinflations. This is yet another claim that is unsupported by any effective proof.

5) *The Choice between the Rate of Change of Global Expenditure and of Prices.* Scadding asserts that no significance attaches to the choice of the rate of increase of global expenditure rather than the rate of price increase p. There is no basis for this assertion, which comes down to saying that if Z is defined by the relation $V/V_0=(1+e^{\alpha Z})/2$, the verification of equation (3) is not affected if x is replaced by p, or in other words that it is always possible to find a value of ϕ_0 such that the fit is equally good whether x or p is taken. This is a wholly gratuitous affirmation, and in fact it is invalidated by calculations undertaken on an alternative basis.

6) *The Use of the Fourier Transform.* Finally, Scadding puts forward an argument in which, assuming that the derivatives of x are bounded, he deduces the existence of an underlying identity, and so is not surprised that ". . . the Allais estimates are very good approximations" (p. 154, fn. 5). This simple assumption, for which no justification whatever is presented,[18] is the starting point for a chain of reasoning based on the Fourier transform which is, to say the least, dubious if not incomprehensible[19] and attempts to cast doubt on the economic meaning of the hereditary and relativistic theory.

---

[18] There is no evidence that the successive derivatives of x for the series United States 1918–41 remain bounded; rather the opposite is true.

[19] I have read and reread Scadding's fn. 5, and consulted pages 32–34 of Papoulis' work to which he refers. Leaving aside the fact that the reference to relation (15) is erroneous and should no doubt be to relation (5) of his text, I cannot see how the properties of the Fourier transform can apply to z. Perhaps Scadding might feel called upon to have a detailed and convincing note on this question published in the *Review*. I have my doubts that he can, for the proposition he tries to demonstrate is really undemonstrable unless the hypothesis is made first that $V=V^*$, which precisely presupposes the validity of the hereditary and relativistic theory (see Section III.2) and III.3) above). In fact, this assumption, which is necessary for his reasoning to be consistent, has nothing to do with the Fourier transform.

REFERENCES

M. Allais, "Reformulation de la Théorie Quantitative de la Monnaie," *SEDEIS Bulletin*, no. 928, Sept. 10, 1965, suppl., Paris.

———, "A Restatement of the Quantity Theory of Money," *Amer. Econ. Rev.*, Dec. 1966, *56*, 1123–57.

———, (1969a) "Growth and Inflation," *J. Money, Credit, Banking*, Aug. 1969, *1*, 355–426.

———, (1969b) "Growth and Inflation, A Reply to the Observations of the Discussants," *J. Money, Credit, Banking*, Aug. 1969, *1*, 441–62.

———, "Allais' Restatement of the Quantity Theory: Reply," *Amer. Econ. Rev.*, June 1970, *60*, 447–56.

———, "Forgetfulness and Interest," *J. Money, Credit, Banking*, Feb. 1972, Part I, *4*, 40–73.

———, "The Psychological Rate of Interest," *J. Money, Credit, Banking*, Aug. 1974, *3*, 285–331.

A. Papoulis, *The Fourier Integral and Its Applications*, New York 1962.

J. L. Scadding, "Allais' Restatement of the Quantity Theory of Money: Note," *Amer. Econ. Rev.*, Mar. 1972, *62*, 151–54.

Professor Allais' Theory of the Demand for Money: Rejoinder

By JOHN L. SCADDING*

I find myself in a situation like that of Moliere's M. Jourdain who was surprised to learn that he spoke prose. I am surprised to learn that—according to Maurice Allais, at least—I commit paralogisms.

If all I wanted to do was dispute that, I would not take up space with this rejoinder; the law of diminishing returns applies with special force to these running controversies. The problem is, however, that Allais does not address himself directly to my criticism, and I think the criticism is important because it is fundamental.

My point was a methodological one: that the tests of Allais' theory are probably quite weak. Allais' skirting of this point is unfortunate for two reasons. The first is that one might get the idea, from reading his reply, that his hereditary and relativistic formulation of the demand for money was under attack. The second is that one might think that he provides confirmation of his theory when I think that a careful scrutiny of his test procedures suggests otherwise.

On the first point, nothing was further from my mind when I wrote my comment than an attack on the *ideas* in the hereditary and relativistic formulation of the demand for money. Quite the opposite was the case. I had then, and I continue to have, nothing but admiration for Allais' theory, which I think is original and may ultimately prove fruitful.

I addressed myself only to the second point. The nub of the problem is that, in going from theoretical to empirical specification, Allais introduced an approximation which very likely robbed his tests of much power. That approximation consists of using velocity as a measure of the rate of forgetfulness (1966, p. 1135, equations (2.38) and (2.41)). That means that the estimated coefficient of psychological expansion[1] is a function, among other things, of past velocity. Since that coefficient is used to predict velocity, Allais' testing comes down to estimating velocity as function of its past. It could come perilously close, in other words, to estimating velocity as an autoregressive process. And if that is true, the test of the structural content of the theory is minimal.

Allais' characterization of the above argument is that I accuse him of circularity. That is simply not true. What I suggest is that his tests have little power against the naive alternative of an autoregressive specification for velocity. It may be that velocity *is* a good measure of the rate of forgetfulness. But I have serious doubts as to whether we are going to be able to test that—or the other parts of the theory—by appealing to the behavior of velocity.

Allais in fact concedes that his test must be weak, although he does not say so explicitly. He grants that my argument is correct if we can identify measured with desired velocity (what he calls observed and estimated velocity): "... Scadding unfortunately fails to make the distinction between the observed and estimated values V and V^*. In fact, his mathematical reasonings are valid *only if* V *and* v *are replaced therein by* V^* *and* v^*" (p. 456).

Yet in his original paper, Allais makes the same assumption in deriving the empirical form of his hypothesis: "... but it can reasonably be suggested that the discrepancy between the actual and the desired value of money holdings is always relatively small. ... It further follows that it is possible to write as a first approximation ... (3.2) ... (3.3) $\phi_D \approx \phi$ ... " (1966, p. 1138).

---

* Stanford University.

[1] The analogue, say, of Cagan's anticipated rate of inflation. I use the word analogue because Allais takes exception to thinking of his psychological variables as in any sense anticipated or expected variables.

But ϕ_D and ϕ are the inverses of V^* and V. Allais therefore assumes for empirical testing that V^* and V *are* the same. And he concedes that my analysis is correct under those circumstances. So as far as I can see, we are in complete agreement.

REFERENCES

M. Allais, "The Hereditary and Relativistic Formulation of the Demand for Money: Circular Reasoning or a Real Structural Relation?" *Amer. Econ. Rev.*, Mar. 1975, *65*, 454–64.

———, "A Restatement of the Quantity Theory of Money," *Amer. Econ. Rev.*, Dec. 1966, *56*, 1124–57.

J. L. Scadding, "Allais' Restatement of the Quantity Theory of Money: Note," *Amer. Econ. Rev.*, Mar. 1972, *62*, 151–54.

Required Disclosure and the Stock Market: Comment

By IRWIN FRIEND AND RANDOLPH WESTERFIELD*

In his recent article in this *Review*, George Benston concludes that empirical analysis provides no support for the belief that the disclosure and related provisions of the Securities and Exchange Act of 1934 were either needed or desirable. Since Benston considers disclosure the basic reason for enactment of the 1934 Act, he is highly dubious about the usefulness of this legislation. We shall indicate that the conclusions he draws from his analysis seem faulty.

It should be pointed out at the outset that Benston covers only the Securities and Exchange Act of 1934 and not the Securities Act of 1933. At times, however, his conclusions do not seem to be limited solely to the particular disclosure requirements of the 1934 Act, but instead pertain to both acts. For example, the last sentence in his paper states, "Certainly there is doubt that more required disclosure is warranted" (p. 153), which does not necessarily follow from his previous conclusion that the 1934 Act had no effect on outstanding issues traded on the New York Stock Exchange (*NYSE*).

It is important to note, therefore, that while disclosure is the central feature of 1933 Act, it is only one, though an important, feature of the 1934 Act. Since the increase in disclosure effected by 1933 Act is very much greater than that effected by the 1934 Act, we think that the new issues evidence of the 1933 Act is highly relevant to the likely impact of the disclosure requirements in the 1934 Act. Thus both the 1933 Act and the 1934 Act must be included for a prudent assessment of the efficacy of disclosure requirements. We interpret the evidence on the 1933 Act as clearly favorable to disclosure.[1]

Our own interpretation of the evidence other than Benston's directly relevant to the evaluation of the 1934 Act is that it supplies some support to the thesis that this legislation may have improved efficiency in the market for outstanding stock, but that the evidence here is not so strong as for new issues (see Friend (1972), (1969)). The rest of this comment will examine the new evidence developed by Benston to analyze the impact of the 1934 Act disclosure, with special reference to his two major sets of tests.

I. Impact of Changes in Accounting Data on Common Stock Prices

The first of these tests, which was presented at greater length in earlier studies (see Benston (1967), (1969a)), estimates cross-sectional regressions in the year 1964 between changes in prices of individual stocks on the New York Stock Exchange (adjusted for movements in the market) and "unexpected" changes in each of a number of different financial variables (net sales, cash flow, net operating income, and adjusted net income), with changes in dividends and industry dummies included in all of these relationships. Expected changes in these financial variables were derived from several simple autoregressive models based on past data and then compared with subsequent published data to obtain estimates of unexpected changes. A similar but more limited analysis was carried out for 1963.

For reasons which are obscure, Benston does not use more than one of these unexpected changes in financial variables in the same regression, but only one at a time. Even so, his findings suggest statistically significant relations between price changes and the unexpected changes in each of these financial variables. Nevertheless, in view of what he considers the small size of the regression coefficients, he concludes that ". . . this evidence is not consistent with the underlying assumption . . . that the financial data made public are timely or relevant, on average" (p. 139).[2] There does not seem to be any justification for his willingness to dismiss out of hand the economic importance of these statistically significant results. Thus, he in effect considers not too relevant for stock prices knowledge about changes in financial variables, in spite of the fact that he finds an increase of 100 percent in the annual rate of net sales is associated with an increase in price of 10.4 percent in the month of the announcement, and that changes in other variables are also associated with significant though proportionally smaller changes in price.[3]

---

* Professor of finance and associate professor of finance, respectively, University of Pennsylvania.

[1] In addition to the papers cited by Benston, see also Friend and Edward Herman.

Moreover, it seems clear that Benston's regressions considerably understate the usefulness of published financial statements. They do not allow for the joint effects of unexpected changes in the different financial variables, and they make no adjustment for the substantial understatement of the relevant regression coefficients arising from the very large random measurement errors associated with any empirical measures of unexpected change.[4]

II. Effects of Disclosure on Expected Return and Risk

Benston's second set of tests, presented for the first time in his 1973 paper, is intended to determine whether stocks which were affected by disclosure provisions of the 1934 Act behaved "better" relative to the market as a whole than stocks for which disclosure was already available prior to the Act. For purposes of these tests, he classifies 508 *NYSE* stocks in the predisclosure period into two groups, one consisting of 193 stocks which did not disclose their sales versus 314 which did. Benston uses his earlier study to justify the selection of sales, maintaining it indicates that sales was the only relatively important accounting number. However, net income would seem to be the more theoretically relevant variable and in fact is the more consistently significant in its effect on price according to his earlier test ((1967), (1969a)), even though the estimated elasticity of price with respect to income was considerably smaller than with respect to sales.

It is interesting therefore that (with one possible exception) all of the 193 stocks which did not disclose sales did disclose net income as well as balance sheet and other financial data, so that Benston's second set of tests does not distinguish between nondisclosure and disclosure firms but between less and more disclosure firms that trade on the New York Stock Exchange. Moreover, this type of test tells us nothing about the relative quality of disclosure for both groups of firms before and after the 1934 Act, so that with relatively moderate differences in the quantity of financial data disclosed for these two groups, changes in the quality of disclosure which are likely to affect both disclosure and nondisclosure firms similarly may well be much more important.

To distinguish between the relative performance of nondisclosure and disclosure firms after the 1934 Act, Benston estimates *beta* coefficients, residuals from the market model, and the variance of these residuals for individual stocks in both groups of firms before and after the Act.[5] There is no evidence from the *beta* coefficients that the nondisclosure firms were adjudged relatively less risky in the post-Securities and Exchange Commission (*SEC*) period.[6] This is the only substantive evidence supporting Benston's position. However, even this evidence seems weak to us, because of the question of the difference between the two groups in the degree of nondisclosure. Moreover, there are several external studies showing that current accounting data can be used to make at least as good and possibly superior forecasts of future asset risk (as measured by *beta*) than forecasts dependent only on historically estimated asset risks (see William Beaver, Paul Kettler and Myron Scholes, and Richardson Pettit and Westerfield (1972)).[7] The accounting data which turn out to be useful for this purpose are balance sheet and income account items other than sales.

---

[2] Benston alleges that almost all previous empirical work relating published accounting data to stock price changes leads to the conclusion that the data are not useful or are redundant. However, our limited survey of several recent articles indicates many writers have concluded that published accounting profile variables can be useful in making investment decisions and contribute to market efficiency. See Alvin Martin and Robert May.

[3] Benston does not present a model or a theory that establishes explicit expected values for the regression coefficients in his equation. He states in the preceding article, "If a corporation's financial statements contain unanticipated information, the expectations of the investor about present value of the corporation should change" (1969a, p. 32). Hence, to conclude that disclosure legislation is largely useless because stock prices are not "strongly" affected, especially after statistically significant results are found, seems unwarranted. It puts too heavy a reliance on the intuitive beliefs of the researcher.

[4] The bias introduced from measurement error can be quite large. A discussion of the sources of measurement error when utilizing accounting data is found in William Beaver, Paul Kettler, and Myron Scholes, p. 670–72, which proposes a regrouping procedure to reduce the measurement error. A similar study by Richardson Pettit and Westerfield (1972), e.g., Table 2, pp. 1659–60, indicated that measurement error accounted for 50 to 90 percent of the total variability in the dependent and independent variables in a multivariate regression model. Thus, the estimates of the regression coefficients associated with empirical measures of unexpected changes will probably greatly understate the true values.

Before leaving this discussion of the impact of the 1934 Act on stock *betas*, we should point out that of the three ways the Act might be expected to affect the level and variability of stock prices—through effects on expected return, *beta* coefficients, and residual variability—*beta* coefficients, which measure the covariability of returns on a stock with those on the market, would be expected to be affected least. We have indicated that Benston's analysis suggests that the 1934 Act did improve estimates of expected return. We turn now to its apparent impact on residual variability.

The residuals from the market model in this second set of tests, derived from regressions for each stock fitted to the entire pre-*SEC* and post-*SEC* periods covered, seem to indicate that both disclosure and nondisclosure firms performed better in the post-*SEC* than in the pre-*SEC* periods in the sense that residual risks were reduced for both groups of firms. The algebraic and absolute means of these residuals were computed for disclosure and nondisclosure firms separately for each of four pre-*SEC* periods, for an "adjustment period" (representing the estimated time between congressional action on the bill and "full compliance"), and for three post-*SEC* periods. The absolute means of these residuals, which seem more meaningful than the algebraic means,[8] exhibit a steady increase throughout the pre-*SEC* periods both for disclosure and nondisclosure firms, turn downward in the adjustment period and then continue at the lower level thereafter, with a lower absolute mean in the post-*SEC* than in the pre-*SEC* period.[9]

---

[5] For a recent discussion of the theoretical and empirical properties of the market model, see Marshall Blume.

[6] It should be noted that Benston assumes that an estimate of the *beta* coefficient provides a good measure of the risk of an asset. For this assumption he repeats the "portfolio approach" justification for the *beta* coefficient (see Blume). Contrary to Benston's assertions, none of the assumptions that the portfolio approach depends upon appears strictly correct, and empirical evidence on the validity of the *beta* coefficient is not conclusive. Therefore, it might have been illuminating to have included total variability estimates to supplement the analysis. This is especially true when dealing with individual securities (see Merton Miller and Scholes).

[7] Still another technical problem with Benston's analysis is that he does not indicate how his *beta* coefficients were estimated. We would surmise that he does not adjust for bias in estimating *beta* from contemporaneous data on rates of return on the individual stocks and on the market, and that he implicitly assumes that the universe rate of return on the minimum variance zero *beta* asset is constant over each of the periods covered—in one case the pre-*SEC* and post-*SEC* periods separately, in another the entire period as a whole.

[8] Benston appears mistakenly to consider the sign of the average residual in any time period as a measure of stock performance. He seems to confuse a positive residual with a positive constant term, α_i, in the capital asset pricing line $R_{it}=\alpha_i+R_{0t}(1-\beta_i)+\beta_i R_{mt}+\delta_{it}$, where R_{it} is the one-period return relative of the ith asset, R_{mt} is the one-period return relative of the market portfolio, and R_{0t} is the return relative on a "zero *beta*" asset. The error term δ_{it} is assumed to be independently distributed. The residual terms δ_{it} can be used to measure relative performance only if the intercept terms α_i and the measurement errors in β_i are the same for both sets of firms before and after the disclosure period. Moreover, using the residuals terms to measure relative performance is not strictly compatible with their required properties in estimating the market model.

[9] The absolute means suggest that Benston is mistaken in the first of his two "basic" conclusions that he draws from these data, viz., that " . . . there appears to be somewhat less variance of the residuals of both groups during . . . the adjustment period . . . compared to the following and preceding months" (p. 146).

The pattern of the algebraic means is also revealing. In the four periods before the transition period there is no significant difference between disclosure and nondisclosure firms, although in each period the difference is negative. However, in the first period after the transition period, July 1935–June 1937, the difference between the algebraic means of the residual terms for the disclosure and nondisclosure firms becomes positive and is greater than the differences observed in the other periods.[10] In the subsequent two periods the residuals become negative. Thus, in only one period are the differences in residuals positive. Moreover, this difference appears in the first period after the disclosure legislation (which is the critical period if adjustments and revaluations take place quickly) and does not support Benston's conclusion that the 1934 Act had no effect on stock market prices. In contrast, the results tend to support the hypothesis that managers avoided disclosure to hide their poor performance. The mean of the standard deviations of residuals is also lower in the post-*SEC* period for both groups of firms.

Thus Benston's conclusion that ". . . the '34 Act did not contribute to a reduction in the variance of returns . . . as measured by . . . the standard deviations of the residual returns on securities . . ." (p. 149) may be questioned from his own results (see Table 4, p. 149). It may also be questioned on the basis of a more comprehensive analysis carried out elsewhere (Friend (1972)) which regresses on time the standard deviation of residuals (of the capital asset pricing line) from a series of cross-section relationships of portfolio monthly return and risk for twenty-one periods of twenty-four months each from July 1926 through June 1968. This study finds a statistically significant downward time trend in these residuals. This analysis avoids many of the problems affecting the Benston analysis.

III. Other Analyses

With respect to other arguments Benston makes, he asserts in his Summary and Conclusions: first, that ". . . the effect of the '34 Act on the capital market may have been perverse, since the percentage of corporate expenditures on plant and equipment financed with new public stock and debt issues was lower in the post- than in the pre-*SEC* decades" and second, "With respect to fairness, the evidence that the disclosure requirements did not result in a revaluation of the securities of the affected firms and the conformity of price changes before 1934 to a random walk indicates that the pre-*SEC* stock market was a fair game for investors" (p. 153). So far as the percentage of capital expenditures financed with new issues is concerned, it is not at all clear where Benston draws the justification for his conclusion. The data he presents show that new issues account for a higher proportion of capital expenditures net of depreciation (presumably the only relevant comparison) in 1949–54, the last period he covers, than the corresponding proportion for 1900–24. It is only in the 1920's that a higher ratio is observed, and the size of that ratio (354 percent) might suggest an excessive use of the securities markets in the 1920's or a mistake in arithmetic. Moreover, Benston does not point out as he does in an earlier article (1969b) that the very much higher rate of personal income taxes (relative to capital gains taxes) might be expected to bring about an increase in internal financing in the post-*SEC* period.[11]

---

[10] The student *t*-value for this difference is approximately 1.003 which is clearly not significant under the usual assumptions but is greater in absolute value than the other periods (excepting the transition period).

[11] Another Benston finding that the ranking of industries by the percentage of private placements to total debt issues (which he considers an indication of *SEC*-induced inefficiencies) ". . . was almost exactly the same as a ranking of the extent of bias of the *SEC*'s conservative accounting rules . . ." (p. 151) appears impressive mainly because of the electric and gas, communication, and railroad industries, which have relatively unique methods of debt financing. Another problem with Benston's analysis is the omission of leasing, which can be interpreted as a form of private placement. Railroads are particularly fond of this method of financing. Furthermore, it should be noted that the importance of private placements has markedly decreased since the end of Benston's time period (i.e., 1966).

Benston's last argument relating to the impact of the 1934 Act on fairness is similarly flawed because he draws no distinction between efficiency and fairness in the stock market. While his definition of fairness (p. 152) is not clear, it would appear from his reference to Eugene Fama's "semi-strong form of the martingale hypothesis" (p. 152) that Benston considers a market fair in which corporate insiders make large profits at the expense of other investors so long as this is accomplished quickly.

In fact, none of his tests is relevant to the fairness of the market as between corporate or market insiders and public investors. Yet corporate investors who have access to inside corporate information and stock exchange specialists who have access to inside market data are the only two groups in the market who apparently receive above average returns in the market (see Friend and John de Cani, and Fama). It is our strong impression, based on a reading of congressional reports, scholarly studies, newspaper articles, and many discussions with professionals in the market, that the ability to take advantage of inside information and the relative profitability of corporate insiders and specialists in the pre-*SEC* period were much greater than in the post-*SEC* period.[12] Clearly this would not be reflected in Benston's analysis, which does not distinguish between the experience of insiders and outsiders.

Benston's analysis is relevant (though deficient) to whether disclosure improved certain types of market efficiency, but not to whether disclosure improved fairness as between insiders and the public. This statement applies not only to his specific tests of the impact of disclosure provisions but also to his random walk tests in the pre-*SEC* and post-*SEC* periods. Two further comments on his random walk tests may be in order.

First, random walk tests which examine the degree of serial independence in security returns, such as the runs tests used by Benston, test only the nature of the market reaction to a given set of information or misinformation and do not test for other aspects of market efficiency, namely, the quality and quantity of information produced in the market (see Friend (1972)). Second, a more important market inefficiency occurs when two portfolios of the same risk realize substantially different returns. In a recent paper (Blume and Friend), it was found that in the pre-*SEC* period tested (1928–33) a monthly rebalancing investment strategy applied to a portfolio of *NYSE* stocks (i.e., maintaining initial portfolio weights by buying stocks which declined in price and selling those which rose) yielded very much better, though from a statistical viewpoint not necessarily significantly different, results than a buy and hold policy.[13] This was not true in the post-*SEC* period as a whole. One of several explanations for these results would be an increase in market efficiency from the pre-*SEC* period to the post-*SEC* period due to securities legislation. We feel such a comparison provides a more relevant test of market efficiency than the runs tests used by Benston.

---

[12] For example, see the reports of the *Federal Trade Commission, Securities and Exchange Commission, Special Study of the Securities Markets* and the Senate hearings on *Stock Exchange Practices*.

[13] This was true even after allowance for differential taxes, commissions, and risk.

REFERENCES

W. Beaver, P. Kettler, and M. Scholes, "The Association Between Market Determined and Accounting Determined Risk Measures," *Accounting Rev.*, Oct. 1970, *45*, 654–82.

G. J. **Benston**, "Required Disclosure and the Stock Market: An Evaluation of the Securities and Exchange Act of 1934," *Amer. Econ. Rev.*, Mar. 1973, *63*, 132–55.

———, (1969a) "The Effectiveness and Effects of the SEC's Accounting Disclosure Requirements," in H. G. Manne, ed., *Economic Policy and the Regulation of Corporate Securities*, Washington 1969.

———, (1969b) "The Value of the SECs' Accounting Disclosure Requirements," *Accounting Rev.*, July 1969, *44*, 515–32.

———, "Published Corporate Accounting Data and Stock Prices," *J. Accounting Res., Empirical Research in Accounting: Selected Studies, 1967,* 1–54.

M. E. Blume, "On the Assessment of Risk," *J. Finance,* Mar. 1971, *26,* 1–10.

——— and I. Friend, "Risk, Investment Strategy, and the Long-Run Rates of Return in NYSE Common Stock," *Rev. Econ. Statist.,* Aug. 1974, *56,* 259–69.

E. F. Fama, "Efficient Capital Markets; A Review of Theory and Empirical Work," *J. Finance,* May 1970, *25,* 383–417.

——— and J. MacBeth, "Risk, Return and Equilibrium: Empirical Tests," *J. Polit. Econ.,* May 1973, *81,* 607–36.

I. Friend, "The Economic Consequences of the Stock Market," *Amer. Econ. Rev. Proc.,* May 1972, *62,* 212–19.

———, "The SEC and the Economic Performance of Securities Markets," in H. G. Manne, ed., *Economic Policy and the Regulation of Corporate Securities,*, Washington 1969.

——— and J. de Cani, "Stock Market Experience of Different Investor Groups," *1966 Proc. Amer. Statist. Assn., Bus. and Econ. Statist. Sec.,* Washington 1967, 44–51.

——— and E. S. Herman, "Professor Stigler on Securities Regulation: A Further Comment," *J. Bus., Univ. Chicago,* Jan. 1965, *38,* 106–10.

A. Martin, "An Empirical Test of the Relevance of Accounting Information for Investment Decisions," *J. Accounting Res., Empirical Research in Accounting: Selected Studies 1971,* 1–31.

R. G. May, "The Influence of Quarterly Earnings Announcements on Investor Decisions as Reflected in Common Stock Price Changes," *J. Accounting Res., Empirical Research in Accounting; Selected Studies 1971,* 119–63.

M. H. Miller and M. Scholes, "Rates of Return in Relation to Risk: A Re-examination of Some Recent Findings," in M. C. Jensen, ed., *Studies in the Theory of Capital Markets,* New York 1971.

R. R. Pettit and R. Westerfield, "A Model of Capital Asset Risk," *J. Finance Quant. Anal.,* Mar. 1972, *7,* 1649–68.

——— and ———, "Using the Capital Asset Pricing Model and the Market Model to Predict Security Returns," *J. Finance Quant. Anal.,* Sept. 1974, *9,* 579–605.

U.S. Congress, Senate Committee on Banking and Currency, *Stock Exchange Practices, Hearings,* on S. Res. 84 and S. Res. 56, 73d Cong., Part 17, Washington 1934.

———, *Report of the Federal Trade Commission on Utility Corporations,* 70th Cong., 1st sess., Sen. Doc. 92, Washington, D.C., 1935.

———, *Report of the Securities and Exchange Commission on Investment Trusts and Investment Companies,* Washington 1939–42.

———, *Report of the Special Study of the Securities Markets of the Securities and Exchange Commission,* 88th Cong., 1st sess., House Document 95, 1963, Washington 1963.

Required Disclosure and the Stock Market: Rejoinder

By George J. Benston*

Irwin Friend and Randolph Westerfield's (F-W) comment on my article is appreciated. It certainly is desirable that generally respected legislation, such as the Securities Exchange Act of 1934, be subjected to scientific analysis and that the analysis be subjected to criticism. I shall reply in this rejoinder, item by item, to the critique presented by F-W. For the reader's convenience, the order of their presentation is followed and the most succinct expression of their criticism, wherever possible, is quoted for reference.

1. "We interpret the evidence on the 1933 Act as clearly favorable to disclosure." Since I did not analyze the 1933 Act, the "evidence" does not refer to my paper. Rather their reference is to George Stigler's pioneering analysis as corrected by Friend and Edward Herman (1964, 1965). The data presented in these papers indicate an insignificant difference in the mean rate of return (relative to the market) on stocks floated in the years 1923–27 compared to flotations in the period 1949–55. However, the standard deviations of the returns are higher in the pre-Securities and Exchange Commission (*SEC*) period, which F-W assume is "favorable." Assuming that these findings from an admittedly crude and not well specified "model" can be taken as valid, greater variance of a subset of assets (new issues) should not be considered favorable or unfavorable without consideration of portfolio effects. A constraint by the *SEC* on the availability of assets with a greater variance of returns may prevent investors from constructing portfolios consistent with their preferences for return and risk. Without knowledge of the covariance of returns among securities and other assets and of the effectiveness of the *SEC*, relative to private individuals, in measuring potential risk and in screening out fraudulent issues, one should not ". . . interpret the evidence as *clearly* favorable to disclosure" (emphasis added).

2. ". . . Benston does not use more than one of these unexpected changes in financial variables in the same regression. . . ." This criticism to an earlier, summarized study is correct (though unexpected changes in dividends, quarterly earnings and industry dummy variables were included). However, Nicholas Gonedes (1973) recently completed a similar study which used seven accounting data ratios simultaneously (combined with multiple discriminant analysis into an index): he reports results similar to those of my study.

3. ". . . many writers have concluded that published accounting profile variables can be useful in making investment decisions and contribute to market efficiency" (fn. 2). No such conclusions are drawn in the articles they cite. May (cited in F-W) simply reports a greater than usual variance of price changes in the weeks around the dates when quarterly earnings are announced. He did not posit the conclusion stated by F-W. Martin (cited in F-W) regressed changes in share prices over the period contemporaneous with or prior to the time the accounting statements were published. His study suffers from serious conceptual and statistical problems, which are detailed in the comments published with the article. Moreover, other studies, in addition to those cited in my article, which show no economically significant predictive effect of published accounting data on share prices include Phillip Brown, Gonedes (1971), and William Beaver and Roland Dukes. I will again emphasize here, as I did in my paper, that an economically significant relationship between share

---

* Professor, Graduate School of Management and Center for Research in Government Policy and Business, University of Rochester.

prices (or some other variable) and published financial statements is necessary but not sufficient support for the hypothesis that mandated disclosure is beneficial in some way.

4. "Benston does not present a model or theory that establishes explicit expected values for the regression coefficients in his equation" (fn. 3). The model was presented on pages 3–5 of my 1967 paper (cited by F-W). A similar model (or rationale) is presented in much greater detail in Gonedes (1971) and briefly in Richardson Pettit and Westerfield, p. 1652 (cited in F-W). The signs of the coefficients, though not the magnitudes, are specified or implied in the models described in these papers. It is true, though, that the models do not establish criteria for determining whether a 2 percent change in share price returns coincident with a 100 percent change in the rate of change of reported net income is meaningful or not.

5. ". . . substantial understatement of the relevant regression coefficients arising from the very large random measurement errors. . . ." F-W cite studies which use a *beta* grouping technique to reduce such errors. Gonedes (1973), who used this technique, reports results similar to mine. Further, Westerfield should reread his own statement concerning measurement errors given by Pettit and himself that ". . . errors in measurement of earnings—due in part to accounting inaccuracies— . . . [are] only approximations for the underlying 'true' values" (p. 1660). Given their conclusion, they should question why the *SEC* has done so little to reduce these errors, or even to provide investors with some indication of their magnitude and effect.

6. ". . . net income would seem to be the more theoretically relevant variable [than sales]. . . ." It should be noted that virtually all companies who were affected by the 1934 Act reported net income. Sales was the principal important datum not generally disclosed before the Act and therefore whose disclosure was mandated, de facto, by the Act.

7. ". . . changes in the quality of disclosure which are likely to affect both disclosure and nondisclosure firms similarly may well be much more important." Granted. Unfortunately, neither F-W nor anyone else, to my knowledge, has defined "quality of disclosure" operationally or offered any evidence of greater quality after than before passage of the 1934 Act.

8. ". . . none of the assumptions that the portfolio approach depends on appears strictly correct, and empirical evidence on the validity of the *beta* coefficient is not conclusive. Therefore . . . total variability estimates [should have been included]" (fn. 6). Since assumptions never can be "strictly correct" and empirical evidence can never be "conclusive," I am at a loss to understand the first sentence. With respect to the second sentence, F-W's comment is, I believe, based on the Douglas-Lintner results, given in George Douglas, which report a significant relationship between *ex post* average returns on stocks and the residual variance, rather than total variance. Merton Miller and Myron Scholes (cited by F-W) examined the validity of this result and demonstrated that it could result simply from skewness in the return distributions of the annual data used by Douglas. Michael Jensen and Eugene Fama and James MacBeth also considered this question. Both papers report a substantial relationship between the *ex post* average returns on securities and the systematic risk measured by *beta*. However, they found that the residual variance is not importantly related to the average *ex post* returns of securities over and above that which is already explained by the *beta* coefficients. Jensen concludes: "Thus while the variable may be statistically significant, its *economic* significance is virtually nil . . ." (in footnote 35, p. 367); Fama and MacBeth conclude: "We also cannot reject the hypothesis of the two-parameter model that no measure of risk, in addition to portfolio risk [*beta*], systematically affects average returns" (p. 633). The implication of these studies is that, having considered the *betas*, inclusion of total variability would not "have been illuminating" as F-W assert.

9. "... several external studies showing that current accounting data can be used to make at least as good and possibly superior forecasts of future asset risk...." The Pettit and Westerfield study referred to uses contemporaneous accounting and market data: hence it is irrelevant. The other study referenced to support the quotation, by William Beaver, Paul Kettler, and Scholes, uses accounting data from one decade (1947–56) as instrumental variables to estimate the *betas* of individual securities and portfolios. The accounting data provided very slight improvement in predicting *betas* in the next decade (1957–65) over the naive model (one decade's *beta* is the same as the preceding decade's *beta*) for individual securities and somewhat greater improvement for portfolios. However the magnitude and economic import of the improvement is not given. Gonedes (1973) presents (and I believe better specifies) tests in which he finds statistically significant, but low correlation coefficients (the highest is .45) between market and accounting based estimates of systematic risk. He supposes that the Beaver, Kettler, and Scholes results "... are actually functions of market prices because they used market prices to scale income numbers. Hence, their results may simply reflect 'spurious correlation'" (p. 43). F-W thus somewhat overstate the reported usefulness of accounting data for forecasting *betas*.

10. Benston "... does not indicate how his *beta* coefficients were estimated" (fn. 7). Not so. See pages 142–43. The adjustment period was excluded from the data used to calculate the *betas*. While it may have been preferable to use a portfolio approach to estimate *betas*, the procedure used of comparing disclosure with nondisclosure companies over the same time periods adjusts for "contemporaneous data on rates of return on the individual stocks and on the market" which F-W wish. (These comments also answer fn. 8.)

11. "We have indicated that Benston's analysis suggests that the 1934 Act did improve estimates of expected return." I have been unable to detect where in their comment F-W so indicate. Indeed, F-W do not question the key finding that the residuals (which measure changes in the returns on securities) over the adjustment period are initially the same for companies whose financial reporting were and were not affected by the Act.

12. "The residuals ... indicate that both disclosure and nondisclosure firms performed better in the post-*SEC* than in the pre-*SEC* periods...." First, the reduction in variance did not occur until mid-1933 or mid-1936, depending on how one sees Figures 1 and 2. Second, R. R. Officer concludes that the decrease in share price variability on the New York Stock Exchange was not due to the creation of the *SEC* but to fluctuations in business conditions: "The variability of the market factor in the 1920's and before is similar to postwar variability, and the *SEC* was not formed until 1933" (p. 436). Third, while F-W have the quote given in fn. 9 right, they apparently overlooked the fact that the reference (to months) refers to Figures 1 and 2 rather than to Table 3 which presents two-year averages. Fourth, the relevant test, as indicated in my article, is the difference in the residual variability, etc., of the disclosure and nondisclosure corporations, not in the change in variability of all corporations over time.

13. "Thus Benston's conclusion [based on the standard deviations of the residuals] may be questioned from his own results," and the previous paragraph. I believe a rereading of my paper will show the contrary. The insignificantly greater positive algebraic residuals in the first period after the transition period indicate greater revaluations for the disclosure than for the nondisclosure corporations, if anything. The lower post-*SEC* mean standard deviations of residuals for both groups are not relevant to the hypothesis (see paragraph 12).

14. "It is only in the 1920's that a higher ratio is observed, and the size of that ratio (354 percent) might suggest an excessive use of the securities markets in the 1920's

or a mistake in arithmetic." The ratio was taken correctly from the source cited in the author (1967, p. 68, cited in F-W), where an explanation for the size of the ratio is given. But how does one define "excessive?" Further, are F-W satisfied with corporations' use of capital markets to finance capital expenditures not having returned to pre-*SEC* levels until after 1948?

15. "... ranking of industries by the percentage of private placements to total debt issues ... appears impressive mainly because ... [three of the eight] industries have relatively unique methods of debt financing" (fn. 11). Not so. The rankings of the remaining five industries are in the order predicted with one exception. Nor is the comment that "... the importance of private placements has markedly decreased since [1966]" relevant. It would be surprising had investors and corporations not been able to adjust for the *SEC*'s restrictions on the transmission of information by this date.

16. "... Benston considers a market fair in which corporate insiders make large profits at the expense of other investors so long as this is accomplished quickly." I never said this, though given profitable insider trading (which regulation apparently has not eliminated; see Jeffrey Jaffe), it would be better if whatever information the insiders had was impounded in the market price of shares as quickly as possible. The sooner the information is reflected in the prices, the sooner the price at which outsiders buy and sell is "fair." In any event, I know of no reasoning or evidence that supports the assertion that the amount of insider profits or the probability of their occurrence is affected by required disclosure of corporate financial statements.

17. "It is our strong impression, based on reading of ... [cited material] that ... [insider profits] in the pre-*SEC* period were much greater than in the post-*SEC* period." I have reviewed most of the cited materials and have found no more than anecdotes and assertions. Since specific page references were not provided by F-W, I was unable to check my memory. Apparently we both have formed very different impressions based upon roughly the same "evidence." However, to my knowledge there have been no empirical tests performed which have measured the extent of the insider trading profits in the pre- and post-*SEC* period. Until these measurements occur, F-W and I merely hold different subjective opinions.

18. "In a recent paper (Blume and Friend), it was found that in the pre-*SEC* period tested (1928–33) a monthly rebalancing investment strategy ... yielded very much better, though from a statistical viewpoint not necessarily significantly different, results than a buy and hold policy. This was not true in the post-*SEC* period as a whole. One of several explanations for these results would be an increase in market efficiency ... due to securities legislation." First, I do not understand what "very much better, though from a statistical viewpoint not necessarily significantly different, results" means. Second, Blume and Friend (cited in F-W) offer the explanation quoted as half of the first of three possible explanations. (The other half is "... or because of other changes in the institutional environment" (p. 266).) They continue: "A second explanation may be in the unique character of the stock market decline starting in 1929 and, to a lesser extent, in 1937. A third explanation might attribute the impressive differences ... to chance, reflecting the very large dispersion of the underlying distribution of investment returns. The tests reported in footnote 18 are consistent with this last interpretation" (p. 266). No further discussion follows.

I believe a review of F-W's comment will show that I have covered all of the specific criticisms they make of my paper, including those made in footnotes. Should the reader review my paper, I believe he/she will find evidence and additional substantive empirical tests upon which the conclusion of my paper is based.

REFERENCES

W. H. Beaver and R. Dukes, "Interperiod Tax Allocation, Earnings Expectations and the Behavior of Share Prices," *Accounting Rev.*, Apr. 1972, *47*, 320–32.

P. Brown, "The Impact of the Annual Net Profit Report on the Stock Market," *Australian Accountant*, July 1970, 277–83.

G. W. Douglas, "Risk in the Equity Markets: An Empirical Appraisal of Market Efficiency," *Yale Econ. Essays*, Spring 1969, *9*, 3–45.

E. F. Fama and J. D. MacBeth, "Risk, Return, and Equilibrium: Empirical Tests," *J. Polit. Econ.*, May/June 1973, *81*, 607–36.

I. Friend and E. S. Herman, "The S.E.C. Through a Glass Darkly," *J. Bus., Univ. Chicago*, Oct. 1964, *27*, 382–405.

N. Gonedes, "Some Evidence on Investor Actions and Accounting Messages—Part II," *Accounting Rev.*, July 1971, *46*, 535–51.

———, "Capital Market Equilibrium and Annual Accounting Numbers: Empirical Evidence," unpublished manuscript, Sept. 1973.

J. F. Jaffe, "Special Information and Insider Trading," *J. Bus., Univ. Chicago*, July 1974, *47*, 410–28.

M. Jensen, "Capital Markets: Theory and Evidence," *Bell J. Econ.*, Autumn 1972, *3*, 357–98.

R. R. Officer, "The Variability of the Market Factor of the New York Stock Exchange," *J. Bus., Univ. Chicago*, July 1973, *46*, 434–53.

G. J. Stigler, "Public Regulation of the Securities Markets," *J. Bus., Univ. Chicago*, Apr. 1964, *27*, 117–42.

The Simple Economics of Incentive Contracting: Note

By MICHAEL E. CANES*

In an earlier article in this *Review*, John J. McCall analyzed the effect of certain types of cost incentives on the behavior of firms bidding for government contracts. McCall's major finding was that government sharing in any difference between a firm's bid and actual costs could yield the perverse result that firms whose actual costs are "low" would bid higher than firms whose actual costs are "high," so that the government in selecting the low bidder would choose a high cost firm. Further, the government sharing would result in the high cost firms' submitting cost estimates less than expected costs, so that selection of these firms would result in a tendency toward cost overruns. McCall's findings have important policy implications for they suggest that the government probably has been paying more for its procurements than necessary, that its own policies have been a cause of troublesome cost overruns, and that the type of cost incentive clause analyzed should be excluded from future contracts.

In this note, I shall argue that McCall's findings are the result of a peculiar constraint imposed on the behavior of government contracting officers and of private firms. Further, descriptions of the behavior of contracting officers reported by others seem to contradict McCall's behavioral constraint. Once the constraint is removed, it is shown that the government will choose a "high cost" firm only if it places great weight on a low bid relative to firm willingness to bear deviations of actual costs from bids. Since government contracting officers do not behave as implied by McCall, his conclusions are weakened. Nevertheless, they cannot be entirely dismissed. An alternative policy to that suggested by McCall is proposed herein, and it is shown that the policy will correct the situation while leaving the government the option of sharing in differences between target and actual costs.

I. The McCall Model

Before summarizing McCall's analysis, it is necessary to explain some of the concepts involved. In many procurements and particularly in defense procurement, the government uses a contract type known as "fixed price incentive." In such a contract, a firm provides an estimate of its costs known as a "target cost," and a "target profit" is calculated as a specified proportion of target cost. In addition, any deviation between actual costs and target costs is shared between the government and the firm. Thus, actual costs greater than predicted (cost overruns) reduce firm profits and actual costs less than predicted (cost underruns) increase them, but not by the full amount of the over- or underrun.

The extent of sharing of the difference between target and actual cost is contractually specified. In the extremes, a contractor bears all or none of the deviations of actual from target costs. The former is known as a "firm fixed price" contract, the latter as a "cost-plus fixed fee." This cost sharing arrangement is intended to give contractors incentive to economize while providing that the government bear some of the consequences of deviations between actual and predicted costs. Cost sharing can be beneficial to the government if it can more cheaply bear the consequences of cost overruns than can a contractor.[1]

McCall assumes that firms act to maximize profits by choice of either private sector or government contract opportunities. Profit maximization implies that firms accept government contracts only if their expected profits at least equal those available in the private sector. Under competitive bidding, firm bids for government contracts are set so as to exactly equate expected profits to those available in the private sector.

---

* Assistant professor, Graduate School of Management, University of Rochester.

[1] A contractor can sell off the risk of deviations of actual from predicted costs and hence avoid the consequences. But such sale is itself costly, and presumably the government must compensate a contractor to bear these costs.

McCall first derives his results under assumptions that production costs and private sector receipts are known with certainty, and then proceeds to analyze cases in which these are uncertain. Since no results are significantly changed by relaxing the certainty assumptions whereas simplicity is gained by keeping them, the certainty case is here analyzed.

The notation below follows that of McCall:

C_1 = The actual cost of production, including a return on capital.
C_2 = The cost of the product to the government.
R = Receipts from producing in the private sector.
ρ = Target cost or bid the firm submits.
π_1 = Profits from selling in the private sector.
π_2 = Profits from selling to the government.
α = The target profit rate, a proportion that when multiplied by target costs yields profits; $\alpha > 0$.
$\alpha\rho$ = Target profits.
β = The sharing rate, the fraction of the difference between ρ and C_1 that accrues to the firm; $0 \leq \beta \leq 1$.

From the discussion above, profits derivable from a government contract are given by

$$\pi_2 = \alpha\rho + \beta(\rho - C_1), \qquad \alpha > 0 \qquad 0 \leq \beta \leq 1 \tag{1}$$

and profits in the private sector are[2]

$$\pi_1 = \begin{cases} R - C_1 & \text{when } R > C_1 \\ 0 & \text{when } R < C_1 \end{cases} \tag{2}$$

Under McCall's assumptions, the bid the firm submits on a government contract is calculated by equating π_1 and π_2, and solving for ρ. This yields

$$\rho = \begin{cases} \dfrac{R - (1-\beta)C_1}{\alpha + \beta}, & R > C_1 \\[2ex] \dfrac{\beta C_1}{\alpha + \beta}, & R < C_1 \end{cases} \tag{3}$$

Firms whose private sector receipts exceed production cost are said to be "efficient" or low cost, whereas those whose production costs exceed private sector receipts are "inefficient" or high cost. The important implication of relation (3) is that for an efficient firm, there is a nonpositive relationship between actual costs and bid whereas for an inefficient firm there is a nonnegative relationship.

$$\frac{\partial \rho}{\partial C_1} = \begin{cases} \dfrac{-(1-\beta)}{\alpha + \beta} \leq 0, & R > C_1 \\[2ex] \dfrac{\beta}{\alpha + \beta} \geq 0, & R < C_1 \end{cases} \qquad 0 \leq \beta \leq 1 \tag{4}$$

The crux of the problem is that so long as $0 < \beta < 1$, a firm that expects a cost underrun will bid ***higher*** the *less* his costs of production. In essence, the reason is that with the government sharing in any cost underrun, the less an efficient firm's expected cost, the more it must bid to secure a profit equal to that obtainable in the private sector. This situation can result in inefficient, high cost firms underbidding efficient, low cost ones.

From relation (3), an inefficient firm submits a bid that is below expected costs. Selection of inefficient firms thus results in a tendency towards cost overruns. Also from relation (3), an efficient firm submits a bid below expected cost if $R - C_1 < \alpha C_1$, and above expected cost if $R - C_1 > \alpha C_1$. Thus, selection of efficient firms results in over- or underruns depending upon the sign of the inequality.

Since by assumption the government must pay an efficient firm its private sector alternative and an inefficient firm its costs of production, the relation between government costs and actual costs is

---

[2] McCall assumes that in the long run no firm will endure private sector losses and will always move to another industry in which its profits are zero.

$$C_2 = \begin{cases} R, & C_1 < R \\ C_1, & C_1 > R \end{cases} \tag{5}$$

Awarding of contracts to inefficient firms on the basis of lower bids thus results in higher costs to the government than would be necessary were a different incentive and bidding scheme used.

II. Constraints on the Government and Contractors

McCall's model of contractor behavior is based on premises that α, the profit rate, and β, the sharing rate, are fixed when bids are submitted. This amounts to assuming that government contracting officers are constrained from accepting bids with other than prespecified α and β, and that firms are constrained from offering such bids. To keep the present comment to reasonable length, α will be held fixed by government fiat.[3] But the assumption of a fixed β is inconsistent with government and contractor behavior reported by Frederick Scherer and by John Cross. Both describe negotiations between the government and a single contractor, but their description of government contracting officer behavior provides strong indication of how such officers are likely to view competing bids.

According to Scherer, government contracting officers bargain both for low cost targets *and* high contractor share rates: ". . . generally, the tighter the negotiated cost target is, the higher the sharing proportion desired by the government" (p. 220).

In Cross' view, contractors charge for increases in share rates by raising target costs, and such raising lessens the probability that a cost overrun will occur. According to Cross,

> . . . unexpected contractor cost overruns may appear to reflect upon [procurement officer] performance and to compromise his career . . . [therefore] the government may be expected to *behave* as though it, too, had a strong aversion to risk [i.e., to high government share rates] and, hence, a preference for relatively expensive insurance policies [i.e., for paying a contractor to accept a high share rate]. [p. 211]

Further, "the buyer has always shown some preference for firm fixed price contracts, and, although contract type is usually specified by the government before any bids are taken, a contractor who offers a more stringent type of contract has a considerable advantage over his fellows" (p. 215).

Following Scherer and Cross, it is assumed that government procurement officers prefer high contractor share rates to lower. And following McCall, it is assumed that the government prefers lower cost targets to greater. The implications of making both of these assumptions will now be explored.

III. Implications of Government Preferences for High Sharing Rates and Low Target Costs

Recalling expression (3), for an inefficient firm, a higher share rate dictates a higher bid; that is,

$$\frac{\partial \rho}{\partial \beta} = \frac{\alpha C_1}{(\alpha + \beta)^2} > 0 \tag{6}$$

Thus, an inefficient firm competing for a contract faces a tradeoff between ρ and β. Given the assumptions of the model, such a firm will accept an increased share rate only if it simultaneously increases its target cost. I shall return to this shortly.

For an efficient firm,

$$\frac{\partial \rho}{\partial \beta} = \frac{\alpha C_1 - (R - C_1)}{(\alpha + \beta)^2} \tag{7}$$

Expression (7) is negative if $R - C_1 > \alpha C_1$. In this case, an increase in a firm's share rate allows it to *reduce* its target cost. Thus, an efficient firm with $R - C_1 > \alpha C_1$ can compete for a government contract by simultaneously appealing to government preferences in two dimensions, target cost and share rate. This situation is illustrated by curve I on Figure 1. If $1/\rho$ and β are goods to

---

[3] If firms are allowed to vary α and if they believe it is target profit ($\alpha\rho$) and not the profit rate with which the government is concerned, then for a given target profit, target costs can be set arbitrarily low with profit rate set so as to earn the firm its opportunity cost of capital. Under such conditions, *all* firms will submit bids that are arbitrarily close to zero.

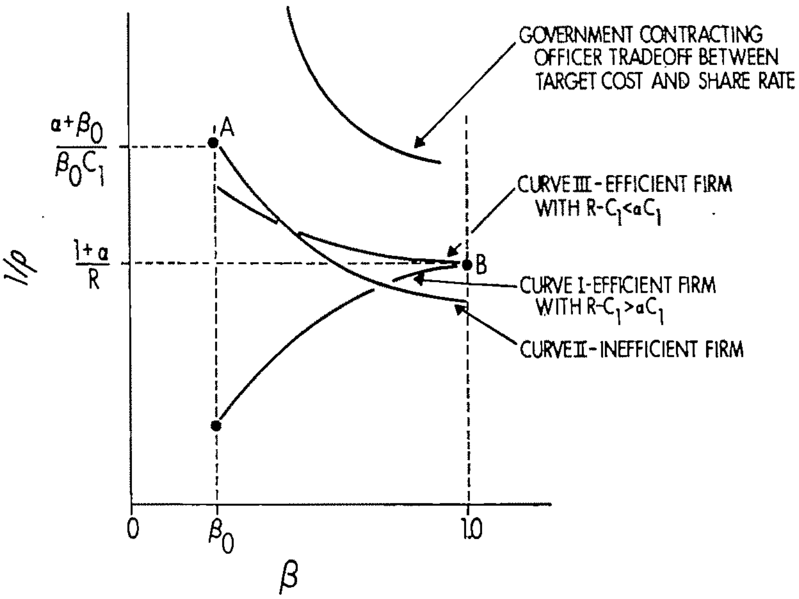


FIGURE 1. BIDS AND SHARE RATES ON GOVERNMENT PROCUREMENT CONTRACTS

government contracting officers, then an efficient firm will move to the northeast along that curve.[4]

Were costs certain and were the share rate not constrained between zero and one, such a firm would reduce target cost to an amount incrementally above its actual cost, with a share rate sufficiently great to cover opportunity costs of accepting the defense contract. However, in actual government contracting the share rate is constrained to be no greater than one. Thus, such a firm bids

$$\rho = \frac{R}{1+\alpha} \qquad (\beta = 1) \tag{8}$$

This is shown as point B in Figure 1.

What does an inefficient firm bid? Apparently, if the firm could obtain a zero share rate (a cost-plus arrangement), it would estimate a zero cost of production. But this is nonsense, for the government would not accept such a bid. Assume that there is some minimum share rate β_0 below which no bids will be accepted. Then an inefficient firm which minimizes its target cost subject to the minimum share rate bids

$$\rho = \frac{\beta_0 C_1}{\alpha + \beta_0} \tag{9}$$

This bid corresponds to point A on Figure 1.

An inefficient firm might choose to increase the share rate it would accept while increasing its bid (i.e., move to the right along curve II). But such a firm's bidding advantage lies in a low target cost-low share rate arrangement, for higher share rates force it to reveal its actual costs. If the inefficient firm has very high costs, then it will be unable to compete with an efficient firm regardless of the share rate it chooses. But over a range of costs, it is possible that an inefficient firm with a low share could submit a lower bid than an efficient firm offering a share rate of one. Such a situation is shown in Figure 1, with point A above point B. In that case, a government contracting officer choosing between an inefficient firm and an efficient firm chooses between a lower target cost, minimum share rate bid and a higher target cost, share-rate-equal-to-one bid. Only if the contracting officer's preference for lower bids outweighs that for higher share rates will the contract go to the inefficient firm.

---

[4] The argument is more complicated if returns from the government project are uncertain, if investors are risk averse in the sense of desiring less variance in a probability distribution over the future returns of their assets, and if there is positive covariance between returns on the government project and those of a market-wide portfolio of assets. In that case, a profit-maximizing efficient firm with $R-C_1>\alpha C_1$ would increase ρ with an increase in β if the market premium for the increased covariance absorbed by the firm were greater than the dollar reduction in government share of a cost underrun. This, however, is not the case analyzed by McCall, and hence will not be followed up here. David Baron has analyzed a somewhat similar problem under the assumption that firms themselves are risk averse.

What of the efficient firm subject to $R-C_1<\alpha C_1$? This inequality is equivalent to expression (7) being positive, so that the firm increases its bid with an increased share rate. If the firm chooses the minimum share rate β_0, then according to expression (7), such a firm bids

$$\rho = \frac{\beta_0 C_1 + (R - C_1)}{\alpha + \beta_0} \tag{10}$$

By expression (9) this bid exceeds that of an inefficient firm by the amount $(R-C_1)/(\alpha+\beta_0)$. But comparing expressions (6) and (7), an efficient firm raises its bid by less with an increase in share rate than does an inefficient firm. Since such a firm is at a competitive disadvantage relative to an inefficient firm at $\beta=\beta_0$, it is to its advantage to compete via higher share rates. Were it to select a share rate of one, then it would bid exactly as does an efficient firm with $R-C_1<\alpha C_1$ (point B on curve III). But it cannot be here determined exactly what such an efficient firm would bid. Presumably that would depend on the firm's beliefs regarding the government's tradeoff between target cost and share rate. The conclusion again is that a government officer confronted with bids from both efficient and inefficient firms will have to choose between a lower target cost with lower share rate, and a higher target cost with higher share rate.

IV. Actual Firm Behavior

Under the assumptions of the model presented, some efficient firms will compete for government contracts by simultaneously increasing the share rate they will accept and reducing their target cost. But Scherer describes contractor behavior as follows: "... the tighter the negotiated cost target is, the lower the sharing proportion a profit maximizing contractor wants" (p. 220). And Cross, p. 210, asserts that contractors charge for higher share rates by raising target costs. This information suggests that only inefficient contractors and efficient contractors with $R-C_1<\alpha C_1$ compete for government procurement contracts.[5] By earlier arguments, these are exactly the class of firms who submit bids below their expected costs. Thus, a systematic tendency towards cost overruns can be expected for the type of contracts here analyzed.

V. A Policy Alternative

From the above discussion, efficient firms can be distinguished from inefficient firms by varying the share rate and choosing the contractor who raises his bid by least. But such a strategy, once discovered by contractors, would surely influence their bidding behavior so as to destroy its advantage.

McCall suggests that the government might set the share rate β equal to one for all procurement contracts. This would induce contractors to reveal their actual costs of production and so would allow the government to select the low cost bidder. But the government would then lose the option of sharing in cost overruns, and under conditions in which the government can more cheaply bear the consequences of an overrun than private contractors, this loss would raise costs to the government.

An alternative would be to set the profit rate α at zero for all cost incentive contracts. Under these conditions, expression (3) collapses to

$$\rho = \begin{cases} \dfrac{\beta C_1 + R - C_1}{\beta}, & R > C_1 \\ C_1, & R < C_1 \end{cases} \tag{3'}$$

According to this expression, inefficient contractors bid exactly their expected costs, independent of the value of β. Further, efficient contractors would be placed in a situation in which they could simultaneously reduce their bid and increase their share rate. Under the assumption that they would set $\beta=1$, (3′) thus collapses to

---

[5] But as noted earlier in fn. 4, under an alternative set of assumptions this behavior does not preclude competition for contracts by efficient contractors with $R-C_1>\alpha C_1$.

$$\rho = \begin{cases} R, & R > C_1 \\ C_1, & R < C_1 \end{cases} \tag{3''}$$

Under these conditions, the government could select the low cost bidder while retaining the option to share deviations between actual and predicted costs. Further, in a situation in which an inefficient firm was the low cost bidder,[6] its selection would result on average in neither a cost over- nor underrun. If an efficient firm were selected, there would result on average an underrun in the sense that actual costs would be below the firm's bid, but neither an overrun nor an underrun in the sense that actual costs to the government would equal its predicted costs. Thus, setting the profit rate equal to zero on fixed price incentive contracts would reduce both expected costs and expected cost overruns on government procurements.

VI. Conclusion

By relaxing an assumption made in an earlier paper by McCall, it has been shown that profit-maximizing behavior by contractors will lead to a different form of competition for government procurement contracts than he suggests. In particular, contractors should be expected to compete both in the bid they submit and the share rate they accept, and inefficient firms will be awarded contracts only if government contracting officers place great weight on a low bid relative to a high share rate. Nevertheless, McCall's principal conclusion that fixed price incentive contracts with share rates between zero and one could result in the government awarding procurement contracts to inefficient firms cannot be dismissed. A suggested policy alternative is to base firm profits only on the deviation of actual from bid costs, i.e., set the target profit rate at zero while allowing firms to choose share rates subject to some minimum rate greater than zero. Such a policy would allow the government to minimize its cost of procurement while inducing firms to reveal their true opportunity costs of producing for the government.

REFERENCES

D. P. Baron, "Incentive Contracts and Competitive Bidding," *Amer. Econ. Rev.*, June 1972, *62*, 384–94.

J. G. Cross, "A Reappraisal of Cost Incentives in Defense Contracts," *Western Econ. J.*, June 1968, *6*, 205–25.

K. L. Deavers and J. J. McCall, *Notes on Incentive Contracting*, The Rand Corporation, RM-5019-PR, Sept. 1966.

I. N. Fisher, *A Reappraisal of Incentive Contracting Experience*, The Rand Corporation, RM-5700-PR, June 1968.

J. J. McCall, "The Simple Economics of Incentive Contracting," *Amer. Econ. Rev.*, Dec. 1970, *60*, 837–46.

F. M. Scherer, (1964a) "The Theory of Contractual Incentives for Cost Reduction," *Quart. J. Econ.*, May 1964, *78*, 257–80.

———, (1964b) *The Weapons Acquisition Process: Economic Incentives*, Cambridge, Mass. 1964.

---

[6] As, say, if only inefficient firms compete.

The Principal Cause of Salary Differentials: Research Output or Experience?: Comment

By WILLIAM HAMOVITCH AND RICHARD D. MORGENSTERN*

David Katz's article in this *Review*, "Faculty Salaries, Promotions, and Productivity at a Large University," which estimates the importance of teaching, research, administrative, and other activities in determining faculty salaries, leaves the reader with an impression that is not supported by the data: namely, that research performance is the single most important determinant of academic salaries. We believe that a careful examination of Katz's own data, as well as the results of other studies, indicate that an equally or more important determinant of faculty salaries is years of experience.

The problem of interpreting Katz's results is really an exercise in stock-flow analysis. Because articles and books published are measured as a lifetime stock, it is necessary to deflate these totals by the years over which they were accumulated in order to get a true picture of the relative importance of these so-called productivity measures compared to experience (years since highest degree) in determining faculty salaries. If this adjustment is not made, the large coefficients on the variables representing publications make it appear that these variables account for the largest part of the differences in salaries among individuals. However, one must recognize that the average individual publishes relatively few books and articles over his career. When all four of Katz's productivity measures (books, articles, excellent articles, and dissertations supervised) are transformed (from his linear equation) and added together, the result is clear: whereas the average year of professional experience is worth $189, the average year's research output is worth $151, or roughly three-quarters as much. If we restrict the productivity variable to the three which include actual publications of books, articles, and excellent articles, the average year's research output is worth $61, or roughly one-third as much as the average year of professional experience. Alternatively, using the partial non-linear estimates presented by Katz for books and articles (which show diminishing returns), the comparable average year's research output is worth $270 when dissertations supervised are included, and $161 when only actual publications are used.[1]

Of the numerous studies that have been conducted on the determinants of faculty salaries, few have actually used such detailed productivity measures as publications. Presumably for lack of data, most have applied a more aggregated analysis and examined only such factors as field of specialty, years of experience, highest degree, and sex. However, several recent studies of individual institutions and of national faculty populations have included measures of research and publications in their analyses of faculty salaries (see John J. Siegfried and Kenneth J. White, Alan E. Bayer and Helen S. Astin, Astin and Bayer, and the authors). When we transformed the findings of these studies in a manner comparable to our transformation of the Katz data, the results showed that by a considerable margin, years of experience had a larger impact in determining salaries on a per year basis

---

* Professor of economics and assistant professor of economics, respectively, Queens College, City University of New York.

[1] The precise calculation for the impact of research, publications, and graduate teaching is the annualized sum of the regression coefficients on the respective variables; that is, each variable coefficient is multiplied by the average value for that variable and then divided by the number of years over which it accrued. The resulting values are then summed. The number of years is 13.13 for all variables except dissertations which is 6.

than did annual research output. In our own study of Queens College, which has a policy of automatic salary increments within each rank, we found that a year's teaching experience was worth almost nine times as much as an average year's research output.

The purpose of this note is to point out the importance of years of experience in determining faculty salaries. Although Katz's study shows research output to be an important factor in salary determination, actual number of years experience appears to be of, at least, approximately equal importance. And, in other studies, experience is found to be by far the most important determinant of faculty salaries.

The obvious question raised by this interpretation is the reason for the high rewards to experience. While the differential rewards for years of service may be due in part to the quality of the overall contribution to the institution, we suspect that a larger part of the salary differential is unrelated to rational market behavior; instead, it probably represents the common practice of rewarding seniority, independent of productivity.

REFERENCES

H. S. Astin and A. E. Bayer, "Sex Discrimination in Academe," *Educational Record*, Spring 1972, *53*, 101–18.

A. E. Bayer and H. S. Astin, "Sex Differences in Academic Rank and Salary Among Science Doctorates in Teaching," *J. Hum. Resources*, Spring 1968, *16*, 192–93.

W. Hamovitch and R. D. Morgenstern, "Sex Discrimination: A Case Study of Faculty at Queens College," in Chancellor's Advisory Committee on the Status of Women at the City University of New York, *The Status of Women at the City University of New York*, New York, Dec. 1972, 225–29.

——— and ———, "The Cost of Being a Women in Academia," *J. New York Affairs*, forthcoming Spring 1975.

D. A. Katz, "Faculty Salaries, Promotions, and Productivity at a Large University," *Amer. Econ. Rev.*, June 1973, *63*, 469–77.

J. J. Siegfried and K. J. White, "Financial Rewards to Research and Teaching: A Case Study of Academic Economists," *Amer. Econ. Rev. Proc.*, May 1973, *63*, 309–15.

The Principal Cause of Salary Differentials: Research Output or Experience?: Reply

By DAVID A. KATZ*

William Hamovitch and Richard Morgenstern (H-M) are certainly correct in saying that experience is at least as important as research output in the determination of salaries. However, a professor will probably never gain any considerable experience at a prestigious university unless he publishes; instead, he will be fired. This is demonstrated by the fact that only 6.4 percent of the associate professors at the time of promotion to that rank had published nothing as compared to 22.4 percent for assistant professors who had not yet been promoted. Only 1.6 percent of the full professors at the time of promotion had published nothing.

The existence of diminishing returns to publishing lends support to H-M's conclusion that experience is possibly the most heavily weighted factor in the reward process. To ignore, though, the very large contribution which publication makes in the salary and promotion process would significantly reduce the predictability of the model. Given the diminishing returns to publishing and the high relative rewards to experience, why do professors continue to publish after gaining tenure? Nonpecuniary rewards—prestige and self-satisfaction—would then explain their behavior. H-M suggest that "a larger part of the salary differential is unrelated to rational market behavior." Actually the opposite may be true. If professors are willing to do research at low pay because of the nonmonetary rewards, why pay them any more than is necessary?

---

* University of Missouri, Rolla.

Segmentation of the Labor Market: Comment

By JOHN M. BARRON*

It has been suggested by Charles Holt in his paper "Improving the Labor Market Trade-Off Between Inflation and Unemployment" (1969) and, in more detail, in the volume edited by Edmund Phelps, that a segmentation of the labor market leads to an increase in the unemployment rate. One purpose of this comment is to clarify the effect of a division of the market for labor services using a model similar to that proposed by Holt. In examining the common notion that combining formerly separated labor markets improves the aggregate performance of an economy, it becomes clear that whether improvement results depends on certain assumptions to be discussed. This result is important in light of the general unqualified feeling that centralized (federalized) manpower programs would reduce structural unemployment (see Holt et al.).

Consider an economy consisting of m firms. Let j_t denote the number of employment positions at the m firms in period t. Define

l = number of people in the labor force (constant)
e_t = number employed in period t
u_t = number unemployed in period t
v_t = number of vacancies in period t

The following relations hold:

$$l = e_t + u_t \quad \text{and} \quad v_t = j_t - e_t$$

To permit a simple presentation, it is assumed that in any given period there is not more than one vacancy per firm.

The model considers a random search procedure by the unemployed workers. In one period, an unemployed worker contacts k firms. A period is defined such that k is sufficiently small ($k<1$) so that the probability of more than one searcher contacting the same firm in a given period approaches zero (a similar assumption is made by Reuben Gronau). For the individual, the average time (number of periods) taken to visit one firm is $1/k$. If there is only one vacancy existing in the economy, the expected number of periods, the "mean search time," before this vacancy is located by a particular searcher is $(1/(1/m))(1/k)$ or m/k.[1] To simplify notation, $\mathrm{T}=(m/k)$ is substituted in the following discussion. Given v_t vacancies, the probability that an individual engaged in a random search process will locate a vacancy in one period is v_t/T.

Let P_e denote the average probability that on finding a vacancy, employment is attained. Thus the probability of employment in period t is $(P_e \cdot v_t)/\mathrm{T}$ for the individual. Aggregating, the total flow of unemployed people into employment, i.e., the flow of hires per period, F, is

$$F = \frac{P_e v_t u_t}{\mathrm{T}} \tag{1}$$

Note that u_t and v_t are stock concepts, F is a flow concept.

Let f_t denote the flow of new entrants into the labor force in the t-th period. It is assumed that new entrants enter the stock of unemployed and proceed to search for vacancies. Let q_t and s_t denote, respectively, the number of people who enter the ranks of the unemployed through quits or layoffs. Then,

$$u_{t+1} - u_t = q_t + s_t + f_t - F \tag{2}$$

Let r_t denote the number of people that retire from employment in period t. Denote the number of new positions created in the t-th period by x_t. Then,

$$v_{t+1} - v_t = q_t + x_t + r_t - F \tag{3}$$

---

* Graduate student, Brown University.

[1] This expected time to locate the vacancy is a random variable having a geometric distribution with parameter (k/m).

In a steady state, $F=q_t+s_t+f_t=q_t+x_t+r_t$. The resulting u_t and v_t may be denoted the equilibrium number of unemployed workers and vacancies. Assuming an equal number entering and leaving the labor force ($f_t=r_t$), maintenance of a steady state requires that the number laid off by certain firms equal the number of new workers required by other firms each period. In the following analysis, it is assumed that the aggregate labor market flows other than hires are constant over time.

Consider an exogenous change, such as an increase in k, signifying an efficiency improvement in the search procedure. The variable T is reduced and F, the flow of hires, increases. Starting from a position of equilibrium, the effect is a negative change in the stock of unemployed workers and vacancies until F has decreased back to its original level. Naturally, this outcome implicitly assumes other things constant. One might, for example, expect P_e to change which may partially offset the above effect. Nevertheless, the implication remains that a reduction in the equilibrium number of unemployed workers results from a more efficient search process.

Now, consider a splitting of the economy into N identical sectors. Holt states that u_t and v_t "would each tend to be increased by a factor of $\sqrt{N}$ as the result of compartmentalization compared to a unified market" (1970, p. 237). A version of his proof follows.

(4) $u_t^i=u_t^d/N=$ number unemployed in the ith sector

(5) $v_t^i=v_t^d/N=$ number of vacancies in the ith sector

where d designates aggregate unemployment and vacancies for a divided economy. Aggregate unemployment u_t^d may differ, as the result of segmentation, from u_t. Employing equation (1),

$$F^i = \frac{P_e u_t^i v_t^i}{\mathrm{T}} \tag{6}$$

where F^i is the flow of hires per period in the ith sector. Defining

$$\sum_{i=1}^{N} F^i = F^d, \quad \sum_{i=1}^{N} u_t^i = u_t^d \tag{7}$$

and
$$\sum_{i=1}^{N} v_t^i = v_t^d$$

and summing equation (6),

$$F^d\mathrm{T} = P_e \sum_{i=1}^{N} u_t^i v_t^i \tag{8}$$

where d again refers to the aggregate value after division of the economy. In the undivided economy, from equation (1),

$$F\mathrm{T} = P_e v_t u_t \tag{9}$$

To maintain the same rate of flow of hires in the divided economy as in the one not divided requires that $u_t v_t = \sum_{i=1}^{N} u_t^i v_t^i$. But, given the above definitions for u_t^i and v_t^i

$$\sum_{i=1}^{N} u_t^i v_t^i = \frac{u_t^d v_t^d}{N} \tag{10}$$

Therefore, to maintain the same aggregate flow F, aggregate unemployment and vacancies must increase due to segmentation. Or, starting at the original aggregate level of unemployment, F falls leading to rising unemployment and vacancies, to the point where u_t and v_t may increase by the factor $\sqrt{N}$.

This conclusion by Holt rests on the assumption that P_e and T are unchanged by dividing the economy. However, his idea of compartmentalization is the isolation of regions. Given a random search procedure, this reduction in the number of firms in each "sub"economy by a divisor of N, so that the number is m/N, directly affects the probability of locating a vacancy. Specifically,

$$\mathrm{T}^i = (m/N)/k \tag{11}$$

This alteration in T is necessary due to the reduced number of firms to search in the ith sector. Returning to equation (6), now

$$F^i = \frac{P_e u_t^i v_t^i}{\mathrm{T}^i} \tag{12}$$

Multiplying by N/N

$$F^i = \frac{P_e u_t^i v_t^d}{\mathrm{T}} \tag{13}$$

or
$$F^i\mathrm{T} = P_e u_t^i v_t^d$$

Following the same procedure as before, equation (1) states that

$$FT = P_e u_t v_t \tag{14}$$

For the economy after segmentation, using equations (7) and (13), we have

$$\sum_{i=1}^{N} \overset{i}{F}T = \overset{d}{F}T = \sum_{i=1}^{N} P_e \overset{i}{u}_t \overset{d}{v}_t \tag{15}$$

But

$$\sum_{i=1}^{N} P_e \overset{i}{u}_t \overset{d}{v}_t = P_e \overset{d}{u}_t \overset{d}{v}_t \tag{16}$$

Note that equation (16) implies no change in the aggregate flow of hires due to compartmentalization. Thus by the above, it is proven that a division of the economy, since the model dictates an adjustment of T for the individual sectors, does not lead to a change in the aggregate level of unemployment or vacancies.

The above appears to contradict the intuition that barriers increase unemployment and vacancies. That intuition may be correct, but it requires a different formulation than given above.

For example, if the market is unknowingly subdivided, such as by discriminatory practices, then T is not reduced, since the individual will approach randomly all apparent employers. What has occurred is that, for those discriminated against, there is de facto a reduction in vacancies, leading to increased unemployment. Or, one might interpret the effect as a reduction in P_e, the average probability of an offer and acceptance of a job, a vacancy having been located. Again, unemployment is increased for this element of the economy—the discriminated worker.

A second implication of the analysis comes by dropping the equal division of unemployed workers and vacancies within equal subdivisions of firms. In the examples that follow, the distribution of vacancies and unemployment in the subeconomies becomes crucial in analyzing the implications. Assume the number of firms in each sector is $1/N$ times the number in the economy; T is reduced by the factor $1/N$ also, as has been seen. But assume the vacancies and unemployed have been separated unequally into the N regions. From equations (11) and (12)

$$\overset{i}{F}T = NP_e \overset{i}{u}_t \overset{i}{v}_t \tag{17}$$

Then

$$\overset{d}{F}T = NP_e \sum_{i=1}^{N} \overset{i}{u}_t \overset{i}{v}_t \tag{18}$$

$$= NP_e(\overset{1}{u}_t \overset{1}{v}_t + \overset{2}{u}_t \overset{2}{v}_t + \dots \overset{N}{u}_t \overset{N}{v}_t)$$

Designate a "pure mismatch" effect as the division of the economy such that in a sector with unemployment above the mean u_t/N, there are vacancies below the mean v_t/N. In this case, it is clear that aggregate unemployment will be higher in the divided economy than if the economy were not split. With a given aggregate number of unemployed and vacancies, the aggregate flow of hires in an economy so divided, as determined by (18), will be less than that of an undivided economy as determined by (9). Designate a "pure match" effect as the division of the market such that in a sector with unemployment greater than or equal to the mean, there are vacancies greater than or equal to the mean. Then it is clear that aggregate unemployment will be below or equal to the level of unemployment before subdivision. One case of equality, discussed previously, occurs when there is an equal division of u_t and v_t among the N sectors. The pure match effect is due to economies of scale in the placement flow. Holt recognized the existence of economies of scale, but failed to accurately state conditions under which they could occur.

A third example is often the one behind intuition concerning the segmentation of markets. Assume that there are Q different nonsubstitutable types of labor. In the extreme case, the market would be divided into Q isolated regions; within each would be the entire unemployed in one category of workers while the vacancies within the region would be those of other categories. In such an extreme, the flow of hires would be zero in the aggregate. This may be likened to an extreme pure mismatch effect, in which in each region the unemployed are mismatched with zero vacancies.

REFERENCES

R. Gronau, "Information and Frictional Unemployment," *Amer. Econ. Rev.*, June 1971, *61*, 290–301.

C. C. Holt, "Improving the Labor Market Trade-off Between Inflation and Unemployment," *Amer. Econ. Rev. Proc.*, May 1969, *59*, 135–46.

———, "How Can the Phillips Curve Be Moved to Reduce Both Inflation and Unemployment?" in E. S. Phelps et al., eds., *Microeconomic Foundations of Employment and Inflation Theory*, New York 1970.

——— et al., "Manpower Proposals for Phase III," *Brookings Papers*, *3*, 703–22, Washington 1971.

Segmentation of the Labor Market: Rejoinder

By CHARLES C. HOLT*

The comment by John Barron constitutes a very interesting extension of earlier work rather than a correction of it. He introduces the additional dimension of searching firms to find vacancies. The earlier work had concentrated on the searching of vacancies to find job offers. Clearly both are relevant and, not surprisingly, he finds that when different assumptions are made, different conclusions follow.

He introduces the concept of mean (firm) search time T to find a particular firm with a vacancy in a given set of firms. Assuming random search, this time depends on the number of firms in the set. Thus, he finds that when a labor market is divided into N equal compartments, mean search time in each compartment is reduced by a factor of $(1/N)$.

The earlier work to which he refers implicitly assumed that the vacancies in a compartment could be located fairly readily. The hiring firms could be easily located but time consuming search was required to determine the particular vacancies which would produce job placements for particular workers. Thus, the relevant search time measure is the mean (vacancy) search time to locate a placement from the set of vacancies in the compartment. This search time depends only on the number of vacancies and is not affected by the number of firms over which the vacancies are scattered. Both concepts of search time are relevant and a comprehensive theory would distinguish between time to search firms for vacancies and time to search vacancies for placements.

Barron finds the aggregated flow relation under his assumption to be

$$F = \sum_{i=1}^{N} \frac{(v/N)(u/N)}{\left(\frac{m}{k}\right) \mathrm{x} \left(\frac{1}{N}\right)}$$

in which the number of compartments N cancels out and has no effect. This cancellation does not occur in the other case in which the mean time to search a vacancy is unaffected by the number of firms in the labor market compartment.

One interpretation of Barron's result is that the search of firms depends only on the probability that a firm will have a vacancy $(v/N)/(m/N)$. Clearly this probability is unaffected by the degree of compartmentalization as reflected in N.

Barron's comment opens up an interesting line of theoretical and empirical work. One essential refinement is to deal with scale effects.

---

* The Urban Institute.

The Elasticity of Scale and the Shape of Average Costs

By GIORA HANOCH*

The theory of production deals with two different concepts of returns to scale. The first, which is most widely used for *defining* returns to scale,[1] is the relative increase in output as all input quantities are increased proportionally along a ray through the origin. The second, which is the more relevant concept for micro-economic analysis, is the increase in output relative to costs for variations along the expansion path where input prices are constant and costs are minimized at every output.[2]

As it turns out, these two concepts yield equal measures for the Elasticity of Scale[3] ϵ at every minimum-cost point in the input space.

However, *the change in ϵ with output* is generally different for the two concepts,[4] since expansion paths do not coincide with rays, unless the production function is *homothetic*.[5] But it is the change of ϵ along the expansion path—and not along the ray—which determines the shape of the average cost curve (AC), and which is therefore relevant for the analysis of firms and industry behavior under competition.[6]

These relations are stated and analyzed below. It is shown that assumptions made in the literature[7] which are related to the behavior of ϵ along rays, or to the slope of the technically optimal surface (where returns to scale are locally constant), are irrelevant—being neither necessary nor sufficient for implying classical U-shaped average cost curves.

The following definitions and results are used here:

(a) A production function $y=f(x_1, \ldots . x_n)=f(x)$ is denoted ***regular*** at x, if it satisfies the following conditions at a (strictly positive) point x:

> $f(x)$ is positive, finite, continuously twice differentiable, strictly monotone and strongly[8] quasi concave at x; i.e., $f(x)>0$, $f_i(x)=\partial f/\partial x_i>0$, and $(-1)^i|B_i|>0$, for $i=1, \ldots, n$ where:
>
> $$B_i=\begin{bmatrix} 0 & f_1 \cdots f_i \\ f_1 & f_{11} \cdots f_{1i} \\ \vdots & \vdots \quad\quad \vdots \\ f_i & f_{1i} \cdots f_{ii} \end{bmatrix}=\begin{bmatrix} 0 & f_x^{i\prime} \\ f_x^i & H_i \end{bmatrix}$$
>
> H_i is the Hessian matrix of order i, and B_i is H_i bordered by the marginal products $f_X^{i\prime}=(f_1, \ldots, f_i)$.

(b) $f(x)$ is denoted *concave regular* at x, if it is regular and strongly concave. That is: $H_i(x)$ is negative definite, satisfying:

---

* Hebrew University and Harvard University. Financial support was received from the National Science Foundation (Grant No. G.S. 39865X). I thank Gary Chamberlin, Zvi Griliches, Ed Leamer, and a referee for comments.

[1] See, e.g., Sune Carlson, Leif Johansen, Ragnar Frisch, F. W. McElroy (1964).

[2] The distinction between these concepts dates back to Francis Edgeworth. But many subsequent writers failed to make this distinction, or to realize its full implications (see fnn. 4, 17, and 20 below).

[3] The Scale Elasticity (see Johansen, p. 64) is known under various names: e.g., "elasticity of production" (W. E. Johnson, p. 507); "passus coefficient" (Frisch, p. 65); "function coefficient" (Carlson, p. 17). Its reciprocal is the elasticity of costs, or the "substitumal cost flexibility" (Frisch, p. 167).

[4] Qualitatively, this difference was pointed out by John Rowe, in his comments on Albert Levenson and Babette Solon, and McElroy (1967). Rowe showed that at a point of maximum marginal returns to scale along the ray, the two variations in ϵ may differ (for two factors).

[5] The term was introduced by Ronald Shephard, to define a monotone transform of a positive homogeneous function. On different definitions of homotheticity and proofs of their equivalence, see Lawrence Lau, and the author.

[6] See Lowell Bassett and Thomas Borcherding, C. E. Ferguson and Thomas Saving, S. Charles Maurice, and Dale Truett and Blaine Roberts, among others, on relations between cost curves, factor prices, and firm and industry size under competition.

[7] E.g., Frisch, p. 171; Truett and Roberts, etc.

[8] "Strongly" means "strictly differentiably." See Peter Newman.

$(-1)^i|H_i|>0$, for $i=1,\ldots,n$. It may be shown, however, that $(-1)^n|H_n|>0$ is sufficient for a regular $f(x)$ to be concave regular at x.[9]

(c) Necessary conditions for minimizing (variable) costs $c=\Sigma p_i x_i$ at $x \gg 0$ under given $p \gg 0$ are:

$$(1) \qquad \lambda f_i(x) = p_i; \qquad i = 1, \ldots, n$$

These conditions are also sufficient if $f(x)$ is regular. The points x satisfying (1) for given p constitute an *Expansion Path.*

(d) By Shephard's Duality Theorem,[10] the (minimum) *cost function* $c(y, p)$, dual to a regular $f(x)$, satisfies the following conditions:

> $c(y, p)$ is finite, positive, continuously twice differentiable, and strictly monotone in (y, p), and linear homogeneous and concave in p. Along the expansion path, $c(y, p)$ satisfies:

$$(2) \qquad \frac{\partial c}{\partial p_i} = x_i; \quad \frac{\partial c}{\partial y} = \lambda; \qquad (i = 1, \ldots, n)$$

Hence $\lambda=\lambda(y, p)$ is marginal cost.

(e) The Average Cost function $AC(y, p)=c/y$ is denoted U-*shaped* (for given p), if it has a unique minimum with respect to y. That is, if $\partial AC/\partial y=0$ for some $y>0$, and $\partial AC/\partial y=0$ implies $\partial^2 AC/\partial y^2>0$. Equivalently, AC is U-shaped if $AC=\lambda$ for some y, and $AC=\lambda$ implies $\partial\lambda/\partial y>0$.

(f) A production function which yields U-shaped AC curves for all $p \gg 0$[11] is denoted *classical.*

I. The Elasticity of Scale and Its Variation with Output

For $f(x)$ regular, the Elasticity of Scale $\epsilon(x)$ is defined as follows:[12]

$$(3) \qquad \epsilon(x) = \left.\frac{\partial \log f(kx)}{\partial \log k}\right|_{k=1} = \frac{\Sigma f_i x_i}{y}$$

Equation (3) is sometimes denoted the "generalized Euler equation," since if $f(x)$ is homogeneous of degree ϵ, satisfying $f(kx)\equiv k^\epsilon f(x)$, then $\epsilon(x)\equiv\epsilon$.

The cost (not necessarily minimal) of producing $f(kx)$ at fixed x and p, is $\bar{c}=k\Sigma p_i x_i$, with $\partial \log \bar{c}/\partial \log k=1$. Therefore:

$$(4) \qquad \epsilon(x) = \left.\frac{\partial \log f(kx)}{\partial \log \bar{c}}\right|_{k=1}$$

providing another interpretation of $\epsilon(x)$ as the elasticity of output with respect to costs along the ray kx, for any arbitrary prices p.

To derive the alternative concept of the elasticity of scale, related to the cost function, substitute equations (1) and (2) into (3):

$$(5) \quad \epsilon(x) = \frac{\Sigma p_i x_i}{\lambda y} = \frac{c}{\lambda y} = \frac{AC}{\lambda}$$
$$= 1 \Big/ \frac{\partial \log c(y, p)}{\partial \log y} = \left.\frac{\partial \log y}{\partial \log c}\right|_p$$

where the last derivative is taken along the expansion path through x. Equation (5) thus establishes the equality of $\epsilon(x)$ for the two concepts.

The change in $\epsilon(x)$ with y along the ray kx (for x constant, at $k=1$) is derived as follows:

$$(6) \quad \left.\frac{\partial \epsilon}{\partial y}\right|_x = \left.\frac{\partial \epsilon(kx)}{\partial k} \Big/ \frac{\partial f(kx)}{\partial k}\right|_{k=1}$$
$$= \frac{1}{f(x)}\left[1 - \frac{\Sigma f_i x_i}{f(x)} + \frac{\Sigma\Sigma f_{ij} x_i x_j}{\Sigma f_i x_i}\right]$$
$$= \frac{1}{y}(1-\epsilon) + \frac{1}{y^2\epsilon} x' H_n x$$

This is generally different from the change in $\epsilon(x)$ along the expansion path through x, as derived below.

---

[9] A sketch of the proof is as follows: If $f(x)$ is regular, then at least $(n-1)$ eigenvalues of H_n are negative (by Sturmian Separation Theorem; see C. R. Rao, p. 52). Since $|H_n|$ equals the product of all n eigenvalues, H_n is negative definite, and f concave regular, if $(-1)^n|H_n|>0$.

[10] See Shephard, W. E. Diewert, or Daniel McFadden, p. 44, for proofs and alternative formulations of duality theorems.

[11] This may be restricted to an open convex cone within the positive orthant in the prices-space. Note that properties (e) and (f) are global, rather than local. The term "classical" was used in Truett and Roberts.

[12] See fnn. 2, 4.

First, using the dual cost function, equations (2) and (5) are used to derive $(\partial\epsilon/\partial y)|_p$ in terms of y and p:

$$\text{(7)} \quad \left.\frac{\partial\epsilon}{\partial y}\right|_p = \frac{\partial}{\partial y}\left(\frac{c}{\lambda y}\right) = \frac{1}{y}(1-\epsilon) - \epsilon\left(\frac{1}{\lambda}\frac{\partial\lambda}{\partial y}\right)$$

where $\lambda = \partial c(y, p)/\partial y$.

Next, $(1/\lambda)(\partial\lambda/\partial y)$ is expressed in terms of x, where x is on the expansion path. Differentiating the identity $f(x) = y$ and equation (1) with respect to y (p constant), and solving for $(1/\lambda)(\partial\lambda/\partial y)$ yields:

$$\text{(8)} \quad \frac{1}{\lambda}\frac{\partial\lambda}{\partial y} = [B_n^{-1}]_{00} = \frac{|H_n(x)|}{|B_n(x)|}$$

where $[B_n^{-1}]_{00}$ is the corner element of the inverse matrix of B_n.

Substitution of (8) into (7) finally gives:

$$\text{(9)} \quad \left.\frac{\partial\epsilon}{\partial y}\right|_p = \frac{1}{f(x)}[1-\epsilon(x)] - \epsilon(x)\frac{|H_n(x)|}{|B_n(x)|} = \frac{1}{y}(1-\epsilon) - \epsilon\frac{|H_n|}{|B_n|}$$

Observe that the second right-hand terms in (6) and (9) are different.

II. Technical Optimality and the Scale Elasticity

A point x corresponds to a firm's long-run equilibrium in a competitive industry with similar firms and free entry if and only if the average cost function has a minimum at x, with respect to variations in y along the expansion path through x. A necessary condition for this in view of equation (5) is that $AC = c/y = \lambda$, or that returns to scale are locally constant: $\epsilon(x) = 1$. In addition, $f(x)$ must be concave regular at x, or $\partial\lambda/\partial y > 0$, where $\epsilon = 1$. Points x which satisfy these conditions are denoted "technically optimal." The following proposition relates the change in ϵ with y at a point of local constant returns, to technical optimality at that point.

PROPOSITION 1: *Let $f(x)$ be regular, and* $\epsilon(x) = 1$.

(i) *$f(x)$ is concave regular, and x technically optimal, if and only if* $(\partial\epsilon/\partial y)|_p < 0$.

(ii) *If* $(\partial\epsilon/\partial y)|_p < 0$ *then* $(\partial\epsilon/\partial y)|_x < 0$; *i.e., ϵ is decreasing along the ray at technically optimal points.*

(iii) $(\partial\epsilon/\partial y)_x < 0$ *is not sufficient for* $(\partial\epsilon/\partial y)|_p < 0$; *i.e., ϵ may be decreasing along the ray (and $\epsilon = 1$), but $f(x)$ not concave, such that $AC(y)$ has a maximum and not a minimum at x, with $(\partial\epsilon/\partial y)|_p > 0$ and x not technically optimal (even though average "ray costs" $\bar{c}/y$ are at a minimum).*[13]

PROOF:

(i) By assumptions, $\epsilon = 1$ and $(-1)^n|B_n| > 0$ (f regular). Therefore, $(\partial\epsilon/\partial y)|_p < 0$ if and only if $(-1)^n|H_n| > 0$ by equation (9). But this is equivalent to f concave regular, and x technically optimal.[14]

(ii) By (i), $f(x)$ is concave regular; hence H_n is negative definite, implying $x'H_n x < 0$. But this implies $(\partial\epsilon/\partial y)|_x < 0$ by equation (6).

(iii) The proof of this part is by example. Consider the function:

$$f(x_1, x_2) = -\tfrac{1}{2}x_1^2 - 2x_2^2 + 3x_1x_2 - 4x_1 + 6x_2 - 2.5$$

at the point $(x_1, x_2) = (1, 2)$. Then $f(1, 2)$ is regular with: $y = 3$; $f_1 = f_2 = 1$; and $|B_2| = 11 > 0$. At prices $p_1 = p_2 = 1$, minimum costs for $y = 3$ is at $(1, 2)$, with: $c = 3$; $AC = \lambda = 1$; and $\epsilon = 1$. But at the point $(1, 2)$:

$$\left.\frac{\partial\epsilon}{\partial y}\right|_x = -5/9 < 0, \quad \text{whereas} \quad \left.\frac{\partial\epsilon}{\partial y}\right|_p = 5/11 > 0$$

Note that in the last example $f(x)$ is concave at $(1, 2)$ along the following directions: The ray (with a slope $(\partial x_2/\partial x_1)|_{x_2=2x_1} = 2$); the tangent to the isoquant (with a slope $(\partial x_2/\partial x_1)|_{y=3} = -1$); constant $x_1 = 1$ and constant $x_2 = 2$ (i.e., $f_{11} < 0$ and $f_{22} < 0$). But it is *convex* along the expansion path (which has a positive slope: $(\partial x_2/\partial x_1)|_p=4/7)$, and along the line of constant returns (which is also positively sloped: $(\partial x_2/\partial x_1)|_{\epsilon=1}=+1)$.[15]

---

[13] See equation (4) above. Clearly, if $(\partial\epsilon/\partial y)|_x < 0$ and $\epsilon(x) = 1$, then $\bar{c}/y = k\Sigma p_i x_i/y$ has a minimum with respect to k at $k = 1$.

[14] See fn. 10.

In view of (e) and (f) in Section I, Proposition 1 may now be applied globally:

COROLLARY 1: *Let* $f(x)$ *be regular, with* x *satisfying* (1). *Then*:

(i) AC *is U-shaped (for given* p*), if and only if* $\epsilon(x)=1$ *for some* $y=f(x)$, *and* $\epsilon=1$ *implies* $(\partial\epsilon/\partial y)|_p<0$. *That is, iff the scale elasticity decreases through* 1 *along the expansion path.*

(ii) $f(x)$ *is classical, iff (i) holds for all* $p\gg 0$. *That is, iff* ϵ *decreases through* 1 *along any expansion path.*

(iii) $\epsilon=1$ *and* $\{\epsilon=1$ *implies* $(\partial\epsilon/\partial y)|_x<0\}$ *are not sufficient for* AC *to be U-shaped, or for* $f(x)$ *to be classical. That is,* ϵ *decreasing through* 1 *along rays does not in general imply U-shaped* AC *curves.*

III. Firm Profit Maximization and the Scale Elasticity

Long-run equilibrium conditions for a competitive *industry* (with free entry and exit and similar firms) require *zero profits*, and therefore technical optimality ($\epsilon=1$). But an individual *firm* facing a given output price p and $f(x)$ regular, may be at equilibrium with profits positive, and therefore $\epsilon<1$.

The conditions for long-run equilibrium for the firm are thus: $\lambda=p$; $\epsilon\leq 1$, and $f(x)$ concave regular: $(1/\lambda)(\partial\lambda/\partial y)=|H_n|/|B_n|>0$. But substitution of these conditions in equations (7) or (9) does not yield a definite sign for $(\partial\epsilon/\partial y)_p$. That is, the elasticity of scale may be either decreasing or increasing along the expansion path at a point of equilibrium of the firm.

On the other hand, if both $\epsilon\leq 1$ and $(\partial\epsilon/\partial y)|_p<0$, equation (9) implies $|H_n|/|B_n|>0$, and thus $f(x)$ concave regular. This is summarized in the following:

PROPOSITION 2: *Decreasing elasticity of scale along the expansion path* $(\partial\epsilon/\partial y)|_p<0$ *at a point where* $\epsilon(x)\leq 1$ *is sufficient, but not necessary, for the point* x *to be a valid profit-maximizing equilibrium point for a competitive firm.*

IV. Classical Production Functions and the Slope of the Technically Optimal Surface

Frisch and others[16] used the following assumption extensively in their analysis (in addition to regularity of f):

ASSUMPTION F (Frisch's "Regular Ultrapassum Law"): The equation $\epsilon(x)=1$ determines an $(n-1)$ dimensional surface which is downward sloping, so that $\epsilon(x)>1$ below it and $\epsilon(x)<1$ above it, along any upward sloping path of inputs. Formally: $(\partial x_i/\partial x_j)|_{\epsilon=1}<0$ for all $i\neq j$; $i, j=1,\ldots,n$. On any path $x(\theta)$ such that $dx/d\theta>0$, $(d\epsilon/d\theta)|_{\epsilon=1}<0$.

Assumption F was invoked in the belief that it was equivalent to $f(x)$ being classical.[17] However, the following proposition proves that it is irrelevant, since it is neither necessary nor sufficient for U-shaped average cost curves.

PROPOSITION 3: *Let* $f(x)$ *be regular. Assumption F is neither necessary nor sufficient for* $f(x)$ *to be classical, with U-shaped* AC *curves.*

PROOF:

(i) To show that Assumption F is *not necessary*, consider the following function:

---

[15] Although this example is only *locally* regular, a similar behavior may be found for a function which is regular for all x, with both factors normal (positively sloped expansion path), provided the expansion path intersects the line $\epsilon(x)=1$ from an *opposite direction*, compared to the ray. The fact that a function which is concave along the rays (globally) is not necessarily concave (unless it is homothetic) was pointed out by Newman and by James Friedman. But the example in Newman (p. 305; taken from R. T. Rockafellar) is not monotone in x_1, x_2; and the example in Friedman is not regular. Another example, with $\epsilon=1$ a downward sloping curve (but x_2 inferior) is in the proof of Proposition 3 (iii) below.

[16] For example, see Frisch, chs. 8, 10; Truett and Roberts.

[17] See Frisch, pp. 170–71; Truett and Roberts, p. 976. It is not clear whether Frisch considers this assumption equivalent, or merely sufficient, for f being classical. Sometimes, normality of all factors is assumed as well (but see his p. 170). (Frisch's "Law" assumes also that ϵ decreases monotonically, and equals zero at higher outputs, but these are irrelevant for the present discussion.)

$f^*(x_1, x_2) = [x_1^{-1/2}+x_2^{-3/2}]^{-1}$. It is easily verified that f^* is regular and classical. Expansion paths are given by the equation: $x_2=(3p_1/p_2)^{2/5}\cdot x_1^{3/5}$, and the constant returns to scale (*CRS*) line where $\epsilon=1$ is: $x_2=x_1^{1/3}$. We see that ϵ decreases through 1 along any expansion path, hence f^* is classical by Corollary 1 (ii). But f^* does not satisfy Assumption F, since the *CRS* line is positively sloped.

(ii) To show that Assumption F is ***not sufficient***, the following example is given:

$$\bar{f}(x_1, x_2) = 15x_1 + 45x_2 - x_1^2 - 16x_2^2 - 12x_1x_2 - 29$$

$\bar{f}$ is regular at the point (1, 1) and its neighborhood. The *CRS* line is negatively sloped:

$$\left.\frac{\partial x_2}{\partial x_1}\right|_{\epsilon=1} = \frac{-2x_1 - 12x_2}{12x_1 + 32x_2} = -7/22 \text{ at } (1, 1)$$

Also, $\epsilon>1$ below and $\epsilon<1$ above this *CRS* line at points near (1, 1); hence Assumption F is satisfied.[18] But the expansion path through (1, 1), corresponding to prices $p_1=p_2=1$ (which minimize costs for output $y=2$ at $x_1=x_2=1$), is negatively sloped (x_2 inferior), and intersects the *CRS* line ***from above***, with a slope equal to: $(\partial x_2/\partial x_1)|_p = -\frac{1}{2}$. Thus *AC* has a ***maximum*** at $y=2$ ($p_1=p_2=1$), and it is not U-shaped.

Clearly, assuming in addition that all factors are ***normal***, with the expansion path positively sloped, would be sufficient for f to be classical. But neither this nor Assumption F is necessary. Any analysis based on these assumptions is therefore limited in its generality, since it assumes away many valid classical functions.

In conclusion, if the production function is ***nonhomothetic***[19] it is the behavior of the scale elasticity ***along the expansion path*** which determines the shape of average costs and the validity of the function for competitive equilibrium. Analysis of the behavior of the elasticity of scale along rays of equiproportional changes in inputs, or along other paths of variation of inputs, does not yield conclusive results, and may lead to errors.[20]

---

[18] In this simple example, Assumption F and regularity of $f(x)$ hold only locally. But similar examples exist with f regular and Assumption F holding globally, for all $x \gg 0$. See fn. 16.

[19] For $f(x)$ homothetic, the scale elasticity is a function of output alone: $\epsilon=\epsilon(y)$. Hence $\partial\epsilon/\partial y$ is also a function of y only, and is independent of the particular direction of change in inputs.

[20] A typical error is to examine the validity of a functional form by examining the behavior of ϵ along rays, and infer from it (explicitly or implicitly) about the shape of average costs. See, e.g., Phoebus Dhrymes and Mordecai Kurz, Alvin Cook. See also McElroy (1967, 1970), Bassett, Levenson and Solon.

REFERENCES

L. Bassett, "Returns to Scale and Cost Curves," *Southern Econ. J.*, Oct. 1969, *36*, 189–90.

——— and T. E. Borcherding, "The Relationship Between Firm Size and Factor Price," *Quart. J. Econ.*, Aug. 1970, *84*, 518–22.

S. Carlson, *A Study on the Pure Theory of Production*, Stockholm 1939.

A. A. Cook, "The Demand for Cooperating Labor in Manufacturing Industries; 1947–1964," unpublished doctoral dissertation, Univ. Washington 1968.

P. F. Dhrymes and M. Kurz, "Technology and Scale in Electricity Generation," *Econometrica*, July 1964, *32*, 287–315.

W. E. Diewert, "An Application of the Shephard Duality Theorem: A Generalized Leontief Production Function," *J. Polit. Econ.*, May/June 1971, *79*, 481–507.

F. Y. Edgeworth, "Contributions to the Theory of Railway Rates," *Econ. J.*, Sept. 1911, *21*, 346–70.

C. E. Ferguson and T. R. Saving, "Long-Run Scale Adjustments of a Perfectly Competitive Firm and Industry," *Amer. Econ. Rev.*, Dec. 1969, *59*, 774–83.

J. W. Friedman, "Concavity of Production Functions and Non-Increasing Returns to Scale," *Econometrica*, Sept. 1973, *41*, 981–84.

R. Frisch, *Theory of Production*, Dordrecht 1965.

G. Hanoch, "Homotheticity in Joint Production," *J. Econ. Theory*, Dec. 1970, *1*, 423–26.

L. Johansen, *Production Functions*, Amsterdam 1972.

W. E. Johnson, "The Pure Theory of Utility Curves," *Econ. J.*, Dec. 1913, *23*, 453–513.

L. J. Lau, "Duality and the Structure of Utility Functions," *J. Econ. Theory*, Dec. 1969, *1*, 374–96.

A. M. Levenson and B. Solon, "Returns to Scale and Spacing of Isoquants," *Amer. Econ. Rev.*, June 1966, *56*, 501–05.

S. C. Maurice, "Long Run Factor Demand in a Perfectly Competitive Industry," *J. Polit. Econ.*, Nov./Dec., 1972, *80*, 1271–79.

F. W. McElroy, "Returns to Scale and Spacing of Isoquants: Comment," *Amer. Econ. Rev.*, Mar. 1967, *57*, 223–24.

———, "Returns to Scale, Euler's Theorem, and the Forms of Production Functions," *Econometrica*, July 1964, *32*, 275–79.

———, "Returns to Scale and Cost Curves: Comment," *Southern Econ. J.*, Oct. 1970, *37*, 227–28.

D. McFadden, "Cost, Revenue and Profit Functions," unpublished paper, Univ. California, Berkeley 1970.

P. Newman, "Some Properties of Concave Functions," *J. Econ. Theory*, Oct. 1969, *1*, 291–314.

C. R. Rao, *Linear Statistical Inference and Its Applications*, 2d ed., New York 1968.

J. W. Rowe, "Returns to Scale and the Spacing of Isoquants: Comment," *Amer. Econ. Rev.*, June 1968, *58*, 548–50.

R. W. Shephard, *Cost and Production Functions*, Princeton 1953.

D. Truett and B. Roberts, "Classical Production Functions, Technical Optimality and Scale Adjustments of the Firm," *Amer. Econ. Rev.*, Dec. 1973, *63*, 975–82.

The Welfare Effects of Spatial Price Discrimination

By William L. Holahan*

In a recent article in this *Review*, M. L. Greenhut and H. Ohta (G-O) demonstrated that a spatial monopolist who adopts a spatially discriminating price would produce a larger output than under a mill price policy. They raise, without elaboration, the possibility that this larger output corresponds to a greater level of net benefits. The purpose of this paper is to examine a model in which the net benefits of both price policies can be measured and to extend the discussion to the case of competition for market areas—called spatial monopolistic competition in location theory (see Martin Beckmann 1968, 1970, 1971).

Section I introduces a model, somewhat simpler than that of G-O, in which the basic result of G-O can be derived and the welfare effects of the two policies can be measured. Here it is found that when there is no significant competition for the firm's market area, spatial price discrimination results in firms producing larger output, serving larger market areas, and promoting greater net benefits than under a mill price policy.

Section II then considers spatial monopolistic competition in which extra normal profits invite invasion of market areas. It is shown that the equilibrium firm output and market area will be smaller under spatial price discrimination than under mill pricing and that if price discrimination is allowed, firms will be forced by free entry to adopt that policy. It is also shown that customers buy more when spatial price discrimination is imposed than under mill pricing, but that they are worse off—greater net benefits are derived under mill pricing.

I. The Basic Model

Consider a simple setting: a linear region of uniform population density v per unit length. Let each customer at a distance r from the firm have a demand given by:

$$q = a - b(P + tr) \tag{1}$$

where q = quantity demanded
P = money price at mill
t = transportation cost per unit distance, and hence
tr = transportation cost facing customers at distance r; and $a, b > 0$.

Also let the cost function of the typical firm be given by:[1]

$$C = mQ + F \tag{2}$$

Let the firm be at the center of a linear market area in which there is a constant density of customers v per unit length. Now equations (1) and (2) can be used to write the expressions for net revenue above fixed cost as a function of r:

$$NR(P, r) = v(P - m)(a - bP - btr) \tag{3}$$

Under price discrimination the object is to maximize $NR(P, r)$ at every distance by solving $\partial NR(P, r)/\partial P = 0$. This price policy is:

$$P(r) = \frac{a}{2b} + \frac{m}{2} - \frac{tr}{2} \tag{4}$$

From (4) we see that the price falls as distance from the firm increases. Substituting (4) into (3) the contribution to net operating revenue as a function of r becomes:

---

* Assistant professor of economics, University of Wisconsin-Milwaukee. I wish to thank Martin Beckmann for helpful comments on an earlier version.

[1] The cost function assumed in (2) above gives rise to economies of scale in production. This is equivalent to assuming that if a firm faces a textbook "U-shaped" long-run average cost (*LAC*) curve it will operate on the downward sloping portion. Beckmann (1968) argues that since costs in a spatial economy equal the sum of production plus transportation costs, "when some cost of plant is fixed . . . economies of scale are always consistent with an optimum scale of plant" (p. 46). In addition, it is often true that an "L-shaped" *LAC* curve is more realistic than the U-shaped curve. For the comparative static results presented here, therefore, the linear total cost curve seems justified.

$$(5)\qquad NR(r:P(r)) = \frac{v}{4b}(a - bm - btr)^2$$

The total profit under price discrimination (πD) is therefore the sum of all these contributions minus fixed cost:

$$(6)\qquad \pi D = 2\int_0^{R_d} \frac{v}{4b}(a - bm - btr)^2 dr - F$$

or

$$(7)\qquad \pi D = \frac{v(a - bm)^2 R_d}{2b} - \frac{v(a - bm)tR_d^2}{2} + \frac{vt^2 bR_d^3}{6} - F$$

where R_d is the radius of the firm's market area, to be derived shortly.

A comparable expression for total profit under mill pricing (πM) can be derived by considering (8):

$$(8)\qquad \pi M = 2\int_0^{R_m} v(P-m)(a-bP-btr)dr - F$$

Equation (8) gives the sum over all locations of the net revenue contributions from sale to customers at each distance r. Here the firm charges the same price to all, regardless of distance. Integrating and maximizing with respect to P gives the profit maximizing mill price:

$$(9)\qquad P = \frac{a}{2b} + \frac{m}{2} - \frac{tR_m}{4}$$

Using (9), equation (8) becomes:

$$(10)\qquad \pi M = \frac{v(a - bm)^2 R_m}{2b} - \frac{v(a - bm)tR_m^2}{2} + \frac{vt^2 bR_m^3}{8} - F$$

The market extremity for a spatial monopolist is that distance at which demand falls to zero due to high total transport costs. Substituting (4) into (1) yields the demand under spatial price discrimination as a function of r:

$$(11)\qquad q_d(r) = \tfrac{1}{2}(a - bm - btr)$$

Setting $q_d(R_d) = 0$ we find the market radius under spatial price discrimination:

$$(12)\qquad R_d = \frac{(a - bm)}{bt}$$

Similarly, substituting (9) into (1) yields demand under mill pricing as a function of r:

$$(13)\qquad q_m(r) = \tfrac{1}{2}(a - bm) + \frac{btR_m}{4} - btr$$

Setting $q_m(R_m) = 0$ we find the market radius under mill pricing:

$$(14)\qquad R_m = \frac{2(a - bm)}{3bt}$$

Profit under the two policies can be compared by substituting (12) and (14) into (7) and (10), respectively:

$$(15)\qquad \pi D(R_d) = \frac{v(a - bm)^3}{6b^2 t} - F > \pi M(R_m) = \frac{4v(a - bm)^3}{27b^2 t} - F$$

Here we see that profit under discriminatory pricing is greater than under mill pricing.

We can also show that the output of the firm is larger under spatial price discrimination than mill pricing. Using (11) we can calculate total output:

$$(16)\qquad Q_d = 2\int_0^{R_d} vq_d(r)dr = v(a - bm)R_d - \frac{vbtR_d^2}{2}$$

Using (12) in (16) we find:

$$(17)\qquad Q_d = \frac{v(a - bm)^2}{bt}$$

Similarly we can calculate total output under mill pricing:

$$(18)\qquad Q_m = 2\int_0^{R_m} vq_m dr = v(a-bm)R_m - \frac{vbtR_m^2}{2}$$

Using (14) in (18) we find

$$(19)\qquad Q_m = \frac{4v(a-bm)^2}{9bt}$$

Comparing (19) and (17) we see that $Q_d > Q_m$. The larger demand from distant customers attracted by the lower spatial price discrimination prices outweighs the smaller demand from nearby customers who pay a higher price under spatial price discrimination. This can be seen more clearly by examining the total price—money plus transportation costs—which customers pay as a function of distance from the production site:

$$(20)\qquad TP = P + tr$$

Using (4) and (9) we can write the expression for TP under spatial price discrimination and mill pricing:

$$(21)\quad TPD = \frac{a}{2b} + \frac{m}{2} + \frac{tr}{2}$$

$$(22)\quad TPM = \left(\frac{a}{2b} + \frac{m}{2} - \frac{tR_m}{4}\right) + tr$$

Here we see that $TPD \lesseqgtr TPM$ accordingly as $r \lesseqgtr R_m/2$. That is, relative to mill pricing, the spatial price discrimination policy imposes a welfare loss on nearby customers in the region $(0, (a-bm)/3bt)$ and results in a welfare gain both on customers in region $((a-bm)/3bt, 2(a-bm)/3bt)$ plus the new customers in region $(2(a-bm/3bt), (a-bm)/bt)$. The price policies are graphed in Figure 1.

Part of what has been shown so far is the basic Greenhut-Ohta result: spatial monopolists employing spatial price discrimination will serve a larger market area, produce a larger output, and make a higher profit than if they employ mill pricing.

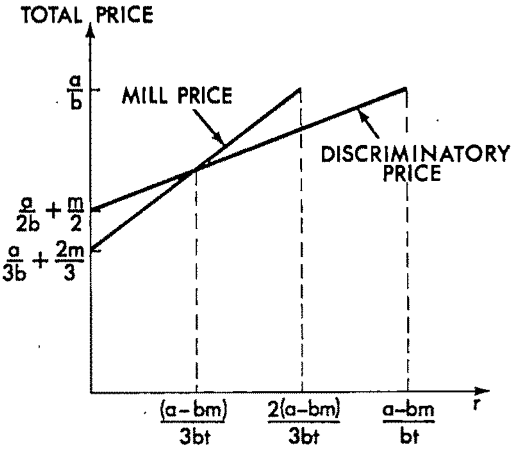


FIGURE 1

We have also seen that there are gainers and losers among the population served, relative to mill pricing. The question arises, how do the two policies compare in terms of a welfare measure? We can address this by using consumers' surplus plus profit as a measure of net benefits. Assuming all customers are alike and using the demand function (1), the consumers' surplus is written:

$$(23)\quad CS = 2\int_0^R \frac{v}{2b}(a - bP - btr)^2 dr$$

Upon substituting (4) into (23) and integrating we get the expression for consumers' surplus under spatial price discrimination:

$$(24)\quad CSD = \frac{v(a-bm)^2 R_d}{4b} - \frac{v(a-bm)tR_d^2}{4} + \frac{vt^2 bR_d^3}{12}$$

Substituting (12) into (24) we get for the discriminatory case:

$$(25)\qquad CSD\left(\frac{a-bm}{bt}\right) = \frac{v(a-bm)^3}{12b^2t}$$

Adding πD and CSD we get the welfare measure under spatial price discrimination

$$(26)\qquad WD = \frac{v(a-bm)^3}{4b^2t} - F$$

Similarly we can compute the net benefits provided under a mill price policy. Substituting (9) into (23) and integrating yields the expression for consumers' surplus under mill pricing:

$$(27)\quad CSM=\frac{v(a-bm)^2R_m}{4b}-\frac{v(a-bm)tR_m^2}{4}+\frac{7vt^2bR_m^3}{48}$$

The level of net benefits provided by mill pricing in the noncompetitive case is the sum of (10) and (27) evaluated at the market radius $R_m=2(a-bm)/3bt$:

$$(28)\qquad WM=\frac{20v(a-bm)^3}{81bt}-F$$

Comparing (26) and (28) we see that spatial price discrimination provides a slightly greater level of benefits. The higher net benefits derived by distant customers under spatial price discrimination, compared to mill pricing, outweigh the loss of nearby customers.

II. The Competitive Case

We can use the equations derived thus far to examine the equilibria arising from the two pricing policies under free entry—invasion of market areas. Letting free entry drive profits to zero, we can use (7) to derive a formula for the equilibrium market size under spatial price discrimination, R_d:

$$(29)\quad \rho\equiv\frac{v(a-bm)^2R_d}{2b}-\frac{v(a-bm)tR_d^2}{2}+\frac{vt^2bR_d^3}{6}=F$$

This is comparable to the equation obtained for the mill price free entry equilibrium market radius R_m found by setting πM in (10) equal to zero:

$$(30)\quad \gamma\equiv\frac{v(a-bm)^2R_m}{2b}-\frac{v(a-bm)tR_m^2}{2}+\frac{vt^2bR_m^3}{8}=F$$

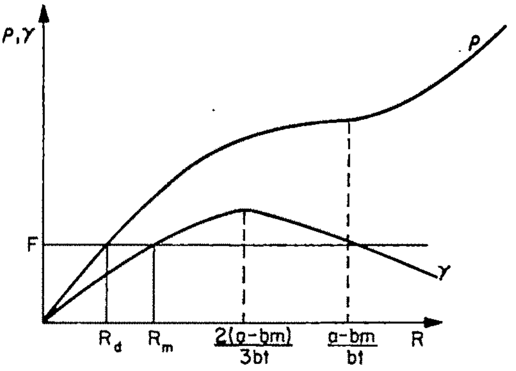


FIGURE 2

Equations (29) and (30) are third degree equations which cannot be readily solved for their roots, but we can show that $R_m>R_d$ by appealing to a geometric argument. Graphing (29) and (30) in Figure 2 we see equilibrium R_m and R_d occur at the intersection of γ and ρ, respectively, and the F line. Since $\gamma<\rho$ for all relevant[2] R, it is clear that $R_m>R_d$. Consulting (16) and (18) we see that $Q_m>Q_d$ since $R_m>R_d$.

It is also clear from (21) and (22) that if spatial price discrimination is allowed, firms will be forced by free entry to adopt it. To see this, let there be a free entry equilibrium under mill pricing. Then by allowing spatial price discrimination we see by (21) and (22) that a firm will be able to undercut a competitor who retains a mill pricing policy at the boundary of their market areas. This will force a competitor to adopt spatial price discrimination in order to offer a matching price at the boundary. Thus the mill pricing equilibrium breaks down. Over time, a new equilibrium will arise in which the region will be served by smaller firms each having a market size given by (29).

---

[2] The distance at which demand falls to zero under the two price policies is the limit of the domain of relevance in Figure 2. As given by (12) and (14) these limits are $R_d=(a-bm)/bt$ and $R_m=2(a-bm)/3bt$. Therefore roots of (29) and (30) which occur beyond those limits are extraneous by economic considerations. Furthermore, for $v(a-bm)^3/6\ b^2t>F>4v(a-bm)^3/27\ b^2t$ firms can survive only if they use the spatial price discrimination policy. We can therefore limit our discussion to $0<R<2(a-bm)/3bt$ and $0<F<4v(a-bm)^3/27\ b^2t$ since these are the ranges in which firms can choose between the price schemes.

We can also see that there will be more output per person under spatial price discrimination because, relative to the mill pricing equilibrium, the increase in the number of firms per area is greater than proportional to the decrease in size of each firm. This can be seen by forming the expression for output per person (QPP):

$$QPP = \frac{Q}{2vR} = \frac{(a - bm)}{2} - \frac{tR}{4} \tag{31}$$

Since $R_m > R_d$ we see from (31) that $QPP_m < QPP_d$.

However, more output per person is not necessarily better. The relevant benefit measure in this spatially extended economy is welfare per area. Dividing (24) and (27) by $2R$, the market area per firm, we can form the expression for consumers' surplus per area under discriminatory pricing ($CSDA$) and mill pricing ($CSMA$), respectively,

$$CSDA = \frac{v(a - bm)^2}{8b} - \frac{v(a - bm)tR_d}{8} + \frac{vt^2bR_d^2}{24} \tag{32}$$

$$CSMA = \frac{v(a - bm)^2}{8b} - \frac{v(a - bm)tR_m}{8} + \frac{7vt^2bR_m^2}{96} \tag{33}$$

Since these equations differ only in the numerical coefficients and not in the parametric coefficients we can normalize the parameters to construct a comparison of welfare per area.[8] Equations (32) and (33) then become

$$CSDA' = \frac{1}{8} - \frac{R_d}{8} + \frac{R_d^2}{24} \tag{32'}$$

$$CSMA' = \frac{1}{8} - \frac{R_m}{8} + \frac{7R_m^2}{96} \tag{33'}$$

Similarly, (29) and (30) can be normalized:

$$\frac{R_d}{2} - \frac{R_d^2}{2} + \frac{R_d^3}{6} = F \tag{29'}$$

$$\frac{R_m}{2} - \frac{R_m^2}{2} + \frac{R_m^3}{8} = F \tag{30'}$$

For every value of F there corresponds a level of net benefits for each price policy. The procedure followed now is to generate the roots of (29′) and (30′) for values of F ranging from zero to 4/27 (the maximum value of F consistent with nonzero profits under mill pricing) and substitute those roots into (32′) and (33′) and compare $CSDA'$ and $CSMA'$. A computer simulation is needed since (29′) and (30′) are third-degree equations and do not yield to algebraic solution.

The roots generated and the level of net benefits as a function of F are given in the Appendix and graphed in Figure 3. It is seen in Figure 3 that mill pricing is superior to

---

[8] By the following manipulations we can see that the result holds for any set of parameter values. Using (29), (30), (32), and (33) note that:

$$CSDA = \frac{F}{4R_d} \quad \text{and,} \tag{a}$$

$$CSMA = \frac{F}{4R_m} + \frac{4vt^2bR_m^2}{96} \tag{b}$$

Hence the inequality: $CSDA \gtrless CSMA$, is equivalent to:

$$\frac{F}{4}\left(\frac{R_m - R_d}{R_mR_d}\right) \gtrless \frac{4vt^2bR_m^2}{96} \tag{c}$$

Now divide (29) by $4R_d$ and (30) by $4R_m$ and subtract and get another expression for the left-hand side of (c):

$$\frac{F}{4}\left(\frac{R_m - R_d}{R_mR_d}\right) = vt^2b\left(\frac{R_d^2}{24} - \frac{R_m^2}{32}\right) - \frac{v(a - bm)t(R_d - R_m)}{8} \tag{d}$$

Using (d) and rearranging, (c) becomes:

$$\left(\frac{R_d^2}{24} - \frac{R_m^2}{32}\right) - \frac{(a - bm)(R_d - R_m)}{8bt} \gtrless \frac{4R_m^2}{96} \tag{e}$$

Since the variables R_m and R_d are measured in distance units of $[(a-bm)/bt]$, it is easily seen that condition (e) is invariant with respect to changes in parameters and hence the welfare comparison in the text holds for all possible parameter sets.

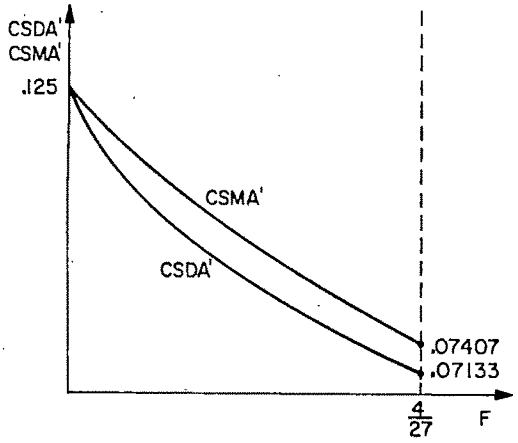


FIGURE 3

spatial price discrimination for all parameter values of F. Therefore, even though the mill pricing policy results in smaller output per person, there is greater benefit per person from mill pricing.

III. Summary

We have found that in regions where competition is insignificant, spatial price discrimination results in larger firms and greater net benefits than a mill price policy. In addition, firms will tend to adopt the better policy since, by using price discrimination to induce distance customers to purchase more, higher profits will be made. Under free entry mill pricing results in larger firms spaced further apart and a smaller output per person. Mill pricing bestows greater net benefits but unregulated firms will be forced by free entry to adopt the worse of the two policies—spatial price discrimination.

APPENDIX

F	R_m	R_d	$CSMA'$	$CSDA'$
.00741	.01504	.01504	.12314	.12313
.01481	.03056	.03055	.12125	.12122
.02222	.04659	.04658	.11933	.11927
.02963	.06319	.06316	.11739	.11727
.03704	.08041	.08036	.11542	.11522
.04444	.09832	.09822	.11342	.11312
.05185	.11699	.11681	.11137	.11097
.05926	.13651	.13621	.10930	.10875
.06667	.15699	.15657	.10717	.10645
.07407	.17863	.17793	.10500	.10408
.08148	.20153	.20046	.10277	.10162
.08889	.22595	.22434	.10048	.09905
.09630	.25218	.24979	.09811	.09638
.10307	.28064	.27709	.09566	.09356
.11111	.31194	.30664	.09310	.09059
.11852	.34700	.33894	.09040	.08742
.12593	.38739	.37474	.08752	.08401
.13333	.43616	.41519	.08435	.08028
.14074	.50124	.46219	.08066	.07613
.14815	.66669	.51923	.07407	.07133

REFERENCES

M. J. Beckmann, *Location Theory*, New York 1968.

———, "Equilibrium Versus Optimum: Spacing of Firms and Patterns of Market Areas," *Northeast Reg. Sci. Rev.*, 1971, *1*, 1–20.

———, "Spatial Cournot Oligopoly," in *Papers of the Regional Science Assn.*, vol. 28, 1972, 37–47.

M. L. Greenhut and H. Ohta, "Monopoly Output Under Alternative Spatial Pricing Techniques," *Amer. Econ. Rev.*, Sept. 1972, *62*, 705–13.

Wage Determination, Inflation, and the Industrial Structure: Comment

By WILLIAM A. BOMBERGER*

In a recent article in this *Review*, Stephen Ross and Michael Wachter (R-W) have argued that in an economy consisting of a competitive sector and a partially noncompetitive sector, there should be a positive relationship between the long-run unemployment rate and the long-run inflation rate if that inflation rate is positive, and a negative relationship if the inflation rate is negative. The purposes of this comment are: 1) to point out that this relationship is not the necessary result of the introduction of noncompetitive elements as is claimed by the authors, but depends instead upon the rather curious use of the planning horizon attributed to the R-W noncompetitive firms; and 2) the model faces serious problems with regard to the existence and attainment of stable equilibrium and the congruence of its predictions with past experience.

R-W construct an economy consisting of two sectors. The competitive sector is characterized by nominal wages and prices which can continuously change and a real wage which is fixed. Firms in the noncompetitive sector are capable of changing wages w^a and prices p^a independently but choose to do so only at regular intervals.

Given the perceived market structure facing the noncompetitive firms a choice of

$$(1) \quad w^a = \phi w^c \quad \text{and} \quad p^a = \mu p^c \quad \mu, \phi > 0$$

would maximize profits at any given time. However, once wages and prices are set, they will not be changed until after a certain time interval has passed. Thus, intervening events may alter the market conditions and produce excess product demand or excess labor supply. R-W assume that customer queuing is more costly than working queuing, and that firms are risk averse. This leads firms to attempt to bring about a market position

$$(2) \quad w^a = \phi\psi w^c \quad \text{and} \quad p^a = \mu\psi p^c \quad \psi > 1$$

which hedges against excess product demand.[1]

Had R-W stopped here, their model would have a short-run inflation-unemployment tradeoff, but a long-run "natural" unemployment rate consistent with any inflation rate. However, the noncompetitive firms are assumed to have the curious habit of being interested only in their market position at some specific time τ within their fixed-price interval.

Clearly if at time T the firm wishes to set wages and prices such that (2) will occur at time $T+\tau$ $(0<\tau<1)$, before it can change wages and prices again at time $T+1$, it must anticipate the change in competitive prices and wages in the interval. If the firm anticipates an inflation rate in p^c of $E(\xi)$, it will set

$$(3) \quad \overset{a}{w}_{T} = \phi\psi \overset{c}{w}_{T} e^{-E(\xi)\tau}; \quad \overset{a}{p}_{T+\tau} = \mu\psi \overset{c}{p}_{T} e^{-E(\xi)\tau}$$

so that

$$(4) \quad \overset{a}{w}_{T+\tau} = \phi\psi E(\overset{c}{w}_{T+\tau}); \quad \overset{a}{p}_{T+\tau} = \mu\psi E(\overset{c}{p}_{T+\tau})$$

If $\tau \geq \frac{1}{2}$, a fully anticipated rise in the inflation rate produces a higher average level of w^a/w^c (averaged, that is, over firms with staggered planning intervals). In the R-W model this is accompanied by a rise in the supply of labor to the noncompetitive sector and an increase in long-run unemployment. In other words, if a firm wants to achieve the proper wage premium at time $T+\tau$, a rise in the inflation rate forces the firm to overshoot the wage premium by a greater amount in the period preceding the relevant date and to suffer a larger excess supply in the labor market until that time.

Despite the claim of R-W this result depends in no way upon either the market power of the noncompetitive firms or the fact that such firms prefer excess supply of labor to excess demand for their products.[2] The reader may simply note that allowing $\psi<1$ (i.e., assuming worker queues to be more costly than customer queues) only shifts the long-run Phillips curve. Its shape is unaltered.

---

* Assistant professor of economics, Wayne State University.

[1] As they state in their article: "Hence, the opportunity loss from underestimating the rate of inflation [creating excess product demand] exceeds that of overestimating [creating excess labor supply], and the risk-averse firm will pay a wage premium. This safety factor is measured by the term $\psi > 1$ . . ." p. 680.

The shape of the long-run curve is the result of the use of the fixed-price interval. If one assumes that such firms change prices and wages only on January 1 each year, for example, R-W assume that wages and prices will be set so as to produce the desired market position on, say, Columbus Day,[3] ignoring the larger excess supply of labor which will precede Columbus Day and the excess product demand which may begin sometime thereafter. A more sensible firm might consider its *average* market position over the year and set a target for the *average* values of w^a/w^c. For the wage rate this means:

$$(5)\quad E(w^a/w^c) = \int_0^1 (w^a_{\mathrm{T}}/w^c_{\mathrm{T}+\tau})\,d\tau = w^a_{\mathrm{T}}/w^c_{\mathrm{T}}(1 - e^{-E(\xi)})/E(\xi)$$

Given the market parameters of the firm and their preference for excess supplies of labor, this means the profit maximizing firm will set

$$(6)\qquad w^a_{\mathrm{T}} = w^c_{\mathrm{T}}(1 - e^{-E(\xi)})^{-1}E(\xi)\phi\psi$$

This changes equation (15) of R-W to

$$(7)\qquad w^a(\mathrm{T})/w^c_{\mathrm{T}} = \phi\psi$$

The average relative wage, and hence the equilibrium unemployment rate, no longer depends upon the inflation rate. The model now exhibits a long-run vertical Phillips curve. The C-shaped relationship was, then, generated by the assumption that workers respond to average wage differentials over the period but firms are interested only in their market position at a specific date within the period.[4]

But the difficulties of the model are not limited to the questionable nature of its key assumption. To illustrate another difficulty let us eliminate the assumption of a target date so that (7) represents a condition of long-run equilibrium. Consider an autonomous increase in demand m. Both sectors will attempt to raise wages and prices and will generate a higher inflation rate. As long as $E(\xi)$ remains unchanged, an equilibrium will be reached at a higher inflation rate characterized by a lower average wage premium offered by the noncompetitive firms which is caused by their underestimation of the inflation rate among competitive firms. This lowers the unemployment rate by shortening the worker queue in the noncompetitive sector. This is illustrated by a movement along the short-run Phillips curve from A to B in Figure 1.

However, as a higher inflation rate persists, $E(\xi)$ must rise. The short-run curve then shifts upward and a new short-run equilibrium is maintained at C. The model is now identical with any natural rate model *à la* Phelps, Friedman, etc. As long as the excess demand is not dissipated by the effects of inflation itself (say a reduction in real balances) or by outside forces, a familiar accelerating inflation rate occurs.

Consider now Figure 2, in which the target date optimization has been reinstated. In such a case there exists a point D at which an equilibrium consistent with the increased demand can be attained. However, the time path of inflation and unemployment is the same as in Figure 1. The equilibrium exists but will not be attained. If such a position were attained, its stability would be questionable but one cannot say more without postulating the relationship between infla-

---

[2] They state: "The emergence of an upward sloping long-run Phillips curve has its roots in the assumption that the administered firm finds a customer queue more costly than a worker queue . . ." p. 688.

[3] In order to generate the proper results it is necessary that the appropriate date be at least halfway through the planning period. If the appropriate date was Valentines Day, we would have a reverse C-shaped relationship with the more traditional Phillips curve in the case of positive inflation and its mirror image in the case of deflation. This fact is apparently recognized by the authors although they fail to make explicit use of it.

[4] It might be reasonable for mathematical convenience to approximate a firm's period-long planning in terms of its view of a certain target date within that period. However no such convenience is realized. The target date approximation is more difficult to handle and leads to perverse firm behavior (see fn. 5 below).

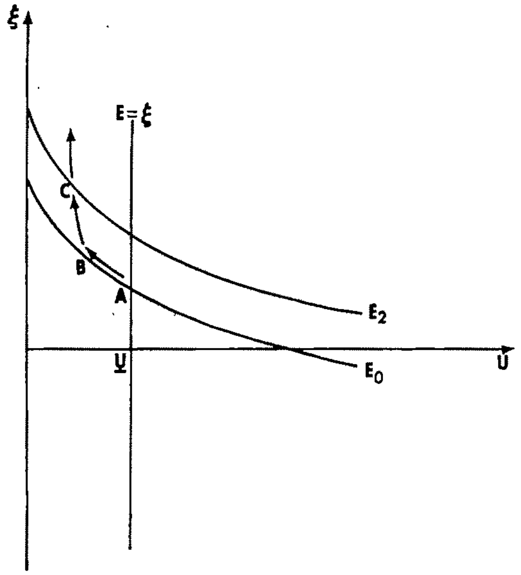


FIGURE 1

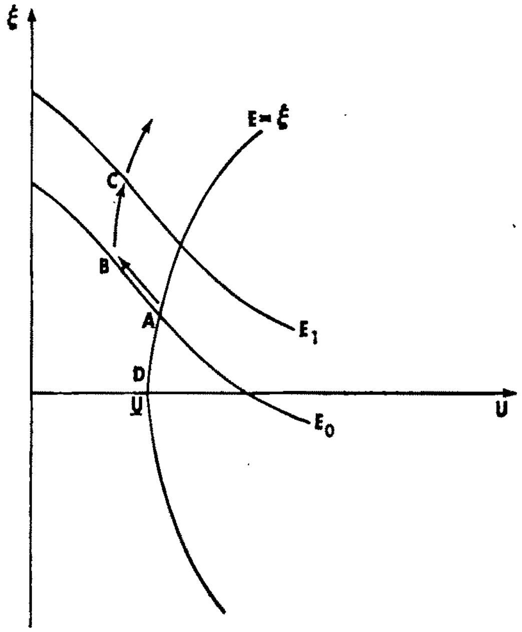


FIGURE 2

tion and excess demand in the competitive sector.

Suppose the R-W economy is in a long-run equilibrium at A with a 5 percent unemployment rate and a 4 percent inflation rate. Suppose further that product demand is increased to such a degree that the new equilibrium unemployment rate at D will be 4 percent. When the economy has again reached equilibrium, the labor queue in the noncompetitive sector must be shorter. Hence, the average wage premium in that sector, w^a/w^c, must be lower. However, according to R-W $\phi'(m)$ is positive—an increased demand for their product induces these firms to raise their wage premium *on the target date.* These two facts can be reconciled if the inflation rate falls sufficiently.

At sufficiently lower rates of inflation the noncompetitive firms will accept a lower average relative wage in order to achieve a higher relative wage on the target date.[5]

To examine the usefulness of this effect assume that in order for equilibrium employment to increase by 1 percent, a 1 percent reduction in the average wage premium is necessary. Further assume that a 1 percent increase in demand convinces the noncompetitive firms to increase their wage premium by 1 percent (ϕ increases by 1 percent). By equation (15) of R-W $h(\xi)$ must fall by 2 percent to achieve equilibrium. Using the R-W formula for $h(\xi)$ (in which Columbus Day falls on July 1, $\tau=\frac{1}{2}$), one finds that ξ must fall to -64 percent to restore equilibrium. Such a result implies a Phillips curve so close to the vertical as to be empirically indistinguishable from it. Except, of course, that one travels the "wrong way" along this curve. The enormous deflation must be generated by an increase in demand. If such a model describes our present economy the recent acceleration of inflation can be attributed to insufficient demand.[6]

REFERENCE

S. A. Ross and M. L. Wachter, "Wage Determination, Inflation, and the Industrial Structure," *Amer. Econ. Rev.*, Sept. 1973, *63*, 675–92.

---

[5] The fact that a desire to increase the relative wage leads to a decrease in its average value is paradoxical. The fact that this lower average relative wage is found acceptable to the firms in this sector illustrates the perverse behavior such a target date "approximation" can generate.

[6] Clearly, different numbers would have produced different results. However, the numbers above are not unreasonable. Unless the labor supply to competitive firms increases significantly in response to truly microscopic changes in wage differentials even as the labor queue in the noncompetitive sector is being shortened, the long-run tradeoff will be extremely steep.

Wage Determination, Inflation, and the Industrial Structure: Reply

By STEPHEN A. ROSS AND MICHAEL L. WACHTER*

William Bomberger in his comment on our recent article raises three points. All are related to our long-run tradeoff between unemployment and inflation. His first point is that our C-shaped Phillips curve is not dependent on noncompetitive forces, as we claim to be the case, but rather is related to our "rather curious use of the planning horizon." His second point is that our long-run model faces serious stability problems. His third point is that our analysis does not yield results which are congruent with the economy's experiences.

Bomberger's first point is incorrect. He seems to believe that the size of the markup variable ψ captures all of the noncompetitive elements in our model. This is not the case. Although the ψ variable is unique to the noncompetitive sector, measuring the safety factor or the risk aversion of the noncompetitive firm, it is not the only factor which distinguishes the noncompetitive sector.

Rather, the main impact of the noncompetitive sector is in introducing the timing or planning period problem. Whereas competitive firms may be viewed as setting wages and prices continuously, the noncompetitive firm acts discontinuously. This creates a planning period over which the noncompetitive firm maintains a fixed wage and price strategy. We also make the important assumption that the firm is more concerned about a customer queue than a labor queue. The result is that it will set a wage premium so as to have a labor queue in the "bulk" of the planning period. At the extreme where it views a customer queue as prohibitive, this will mean setting a wage premium sufficient to insure a labor queue *throughout* the period. The identical result can be obtained by assuming that the elasticity of the supply of labor to the noncompetitive firm is larger than the elasticity of the derived demand for labor. Both assumptions—that the firm is more concerned with a customer than with a labor queue and that the supply elasticity is greater than the demand elasticity—seem to us to be theoretically viable and well established in the empirical literature.

Since Bomberger views ψ as representing the noncompetitive elements in the model, he does not appreciate the role played by the fixed planning period. This leads him to suggest that a more straightforward way of dealing with the planning period is for the firm to consider its *average* market position. This, however, would eliminate one of the basic assumptions of our analysis. Since the noncompetitive firm is more concerned with a customer than a worker queue, it looks toward the *end* of the planning period in setting wages and prices.

Bomberger observes that when firms look to the middle of the period, setting a wage premium sufficient to have a labor queue in the first half of the period and a customer queue in the second, the numerical elasticity of the long-run Phillips curve is quite low. As he verifies in equation (5) by using a period average (ignoring discounting), the wage markup is, in fact, exactly independent of the inflation rate. This is just what one would expect. We analyzed the midpoint case precisely because it gives a lower bound to the slope of the long-run tradeoff. Our point was and still remains, however, that noncompetitive firms aim *beyond* the midpoint.

As for the numerical elasticity, in the extreme where only a labor queue is tolerable, the noncompetitive firm sets $E=e^{\zeta}$ and the aggregate wage premium of our equation (15) is proportional to

$$e^{\zeta} g(\zeta) = \frac{1}{\zeta}(e^{\zeta} - 1) \tag{1}$$

---

* Professor of economics and associate professor of economics, respectively, University of Pennsylvania.

It can be shown that this roughly implies that a 1 percent decline in the inflation rate lowers the relative wage premium by $\frac{1}{2}$ percent. The elasticity of the Phillips curve then depends upon the elasticity of the interindustry labor supply with respect to relative wages. If this elasticity is high, as seems likely to us, then our C-shaped Phillips curve will be quite different from the traditional vertical long-run curve.

Bomberger's second point concerns the disequilibrium behavior of the model. In an earlier version of the paper we developed some of the analytics of this, but the present version contained only the verbal analysis of Section V. We chose not to do formal disequilibrium analysis for two reasons, both of which come to the same thing—it is very difficult to do it right. On the one hand, it requires a sensible mechanism for the formation of expectations in disequilibrium and it did not appear possible to us to easily solve such an important matter. Second, as we observed, a model with an institutional structure will in profound disequilibrium undergo dynamic institutional responses to the inflation rate. One simple and relevant example is the introduction, as inflation increases, of escalator clauses in collective bargaining contracts. This type of institutional response also merits a separate study.

Bomberger's observation that the Phillips curve is unstable, in the sense that movement along the curve in a dynamic context is impossible, should not be surprising. The model does not imply, however, as Bomberger suggests, that an increase in aggregate demand would require a deflation in prices. Suppose that we are initially in equilibrium at a point where a short-run Phillips curve intersects the C-shaped curve. Our disequilibrium analysis (with unchanging expectations and institutions) indicates that an increase in aggregate demand, while initially increasing the inflation rate, will cause the rate to undershoot the short-run Phillips curve and then to move towards that curve. It is likely that the dynamic movement will go beyond the short-run curve, generating a spiraling inflation. That a new static equilibrium may not be reached along the short-run Phillips curve is not disturbing. Since aggregate demand is being fixed beyond full employment, expectations should adjust upward, shifting the short-run curve in the same direction—towards an ever increasing rate of inflation. Observed behavior and the empirical Phillips curve, however, are generated, not by a fixed unemployment rate, but rather by a cyclically changing rate. The dynamic behavior implied by our analysis can generate this type of observed behavior. The short-run downward sloping Phillips curve is a statistical relationship and should not be treated as a static equilibrium relationship.

On the other hand, if one feels that stability is a necessary feature of a dynamic model, then one can add other variables to our analysis to dampen disequilibrium behavior. Our model explicitly abstracts from certain traditional Phillips curve elements in order to concentrate on the impact of introducing a noncompetitive sector with a fixed planning period.

Both our C-shaped Phillips curve and the neoclassical vertical curve are unstable in that any government attempt to reduce and maintain the unemployment rate below its equilibrium level will lead to accelerating inflation. A difference between the two, however, is that in our model, once the government decides to return to full employment, it finds that the equilibrium unemployment rate is higher than it was before the demand-pull inflation was started. (This, of course, assumes that inflationary expectations have adjusted upward.) Whether or not society actually observes this adverse long-run tradeoff as mentioned above, depends upon the nature of institutional adjustments to inflation in the long run. Our own guess is that such an adverse tradeoff might well be expected.

Finally, Bomberger charges that our model does not describe actual behavior. This comment is made in the context of his incorrect supposition that our model requires an increase in aggregate demand to lead to a deflation. If his description of our model were correct, we would agree that it is not in tune with actual events. Fortunately, but not by accident, our model does not make such predictions. We believe that our model does have relevant implications for the inflation problem. In particular, the planning period effect provides a theoretical justification for lags of adjustment longer than that found in neoclassical, competitive models. As a consequence, inflation may not easily respond to government-induced cycles geared to the four-year national election sequence. Furthermore, the distinction between competitive and noncompetitive sectors gives rise to variation in interindustry wages and prices which may appear to be cost-push inflation when, in fact, it is only a lagged response to demand-pull pressure. This traces out a complex pattern of demand-pull inflation where in the initial period of falling unemployment, the competitive industries lead in raising wages and prices. In the second stage, a period of steady or rising unemployment and catch-up wage and price differentials, the noncompetitive industries attempt to reestablish their traditional differentials.

REFERENCES

W. A. Bomberger, "Wage Determination, Inflation, and the Industrial Structure: Comment," *Amer. Econ. Rev.*, June 1975, *65*, 504–06.

S. A. Ross and M. L. Wachter, "Wage Determination, Inflation, and the Industrial Structure," *Amer. Econ. Rev.*, Sept. 1973, *63*, 675–92.

Pitfalls in Financial Model Building: A Clarification

By GARY SMITH[*]

In their well-known "Pitfalls" paper, William Brainard and James Tobin advocated "a 'general disequilibrium' framework for the dynamics of adjustment to a 'general equilibrium' system." The Pitfalls framework has subsequently been widely used in the construction of flow of funds models of financial markets. The Brainard-Tobin paper has also elicited notes from Mark Ladenson and Kevin Clinton in this *Review* which attempt to interpret and elaborate upon the Pitfalls model. However, these authors seriously misinterpret the model and seemingly obscure rather than extend the Pitfalls framework in a maze of unnecessary mathematical techniques and notation.

Ladenson and Clinton both attempt to work with linearly dependent explanatory variables whose coefficients are not identified and cannot be meaningfully interpreted. Clinton uses this indeterminacy to provide a superficial counterexample to a Brainard-Tobin argument. Ladenson, on the other hand, discusses two (of many possible) sets of "expedient" parameter restrictions which will allow him to assign values to all of the coefficients of the linearly dependent variables. He believes that estimation of the model necessitates substantive behavioral assumptions, whereas in fact the elimination of a redundant explanatory variable alters only the appearance of the model and the interpretation of individual coefficients.

In addition to discussing the errors of Clinton and Ladenson, this paper will demonstrate the simplicity with which adding up constraints can be derived, the relationship between the form of a model and the interpretation of its coefficients, and how the parameters might be estimated subject to adding up constraints. In Section I the Pitfalls model is compared with the Clinton-Ladenson formulation; Sections II and III concern the respective details of the Clinton and Ladenson articles; and estimation procedures are discussed in Section IV.

I

All of the authors consider a portfolio model in which desired asset shares depend linearly upon the competing rates of return and other explanatory variables:

$$\frac{a_i^*}{w} = b_{i0} + \sum_{j=1}^{q} b_{ij} r_j \qquad i = 1, \ldots, n \tag{1}$$

where a_i is the actual holding of the ith asset, a_i^* is the desired holding of the ith asset, r_j is the jth rate of return or other explanatory variable, and w is equal to the sum of the a_i.

If, as the authors assume, desired holdings sum to available funds ($\sum_i a_i^* = w$) for all values of the explanatory variables, then

$$1 = \sum_{i=1}^{n} \frac{a_i^*}{w} = \sum_i b_{i0} + \sum_{j=1}^{q} \left(\sum_i b_{ij} \right) r_j$$

Clearly this equation can hold for all values of r_j if and only if

$$\sum_i b_{i0} = 1 \qquad \sum_i b_{ij} = 0 \qquad j \neq 0$$

That is, if all changes in asset value are included in w, then a change in any desired asset share must be compensated by offsetting changes in the remaining shares, and an increase in funds must be fully accompanied by an increase in desired holdings.

All of the authors allow actual holdings to differ from desired shares and assume that holdings will adjust in response to portfolio disequilibria. Brainard and Tobin pointed out that the common specification in which each asset adjusts solely to its own discrepancy between this period's desired and last period's actual holdings

$$\Delta a_i = \epsilon_i [a_i^* - a_i(-1)] \tag{2a}$$

---

[*] Cowles Foundation, Yale University. My research in this area has been supported by grants from the National Science Foundation and the Ford Foundation.

is likely to result in inconsistent or unintentionally implausible behavior since each of these assets is reacting to an incomplete description of the portfolio disequilibria.[1]

The danger of inconsistency arises if specification (2a) is adopted for all n assets. In this case, using $\Delta w = \sum_i [a_i^* - a_i(-1)]$, we have

$$\Delta w = \sum_i \Delta a_i = \sum_i \epsilon_i [a_i^* - a_i(-1)]$$
$$= \epsilon_k \Delta w + \sum_{i \neq k} (\epsilon_i - \epsilon_k)[a_i^* - a_i(-1)]$$

This equation can hold for all potential disequilibria situations if and only if $\epsilon_i = 1$ $i = 1, \ldots, n$. That is, consistency in a pure linear own-adjustment model requires that actual holdings adjust fully so as to always coincide with desired holdings. If some of the ϵ_i are instead allowed to assume values other than 1, then either desired or actual holdings will be inconsistent with available funds for some values of the explanatory variables.

A more frequent practice is to specify own-adjustment equations (2a) for an incomplete list of assets, say for $i = 1, \ldots, q < n$. In this case, the consistency requirement can be used to residually derive the implicit specification of the adjustment equation for the remaining assets as a group,

$$\Delta\left[\sum_{q+1}^{n} a_i\right] = \Delta w - \sum_{1}^{q} \Delta a_i$$
$$= \left[\sum_{q+1}^{n} a_i^* - \sum_{q+1}^{n} a_i(-1)\right]$$
$$+ \sum_{1}^{q} (1 - \epsilon_i)[a_i^* - a_i(-1)]$$

Thus the residual collection of assets implicitly adjusts quite differently from the explicitly specified assets. In particular, the residual collection fully adjusts to its own discrepancy and passively absorbs any funds which are not immediately shifted in accord with desired holdings of any of the assets with explicit adjustment equations. This model could conceivably describe some group's actual behavior. In practice, however, assets are usually included in the residual category not for theoretical reasons, but rather because the model builder is not immediately interested in these items. As a consequence, the specification usually appears to be inadvertant and implausible due to the illiquid nature of the residual assets. One Pitfalls moral is thus that the residual specification should always be directly checked since any inadequacy in this equation must be due to inadequacies in the explicit equations.

To avoid these problems, Brainard and Tobin advocated having each asset adjustment depend upon a complete description of the disequilibria in the portfolio. Ladenson and Clinton instead attempt to work with explanatory variables which overdescribe the portfolio disequilibria:

$$\text{(2b)} \quad \Delta a_i = \sum_{j=1}^{n} \epsilon_{ij}[a_j^* - a_j(-1)] + f_i \Delta w \qquad i = 1, \ldots, n$$

The interpretation of the parameters in this specification is completely arbitrary since the explanatory variables are linearly dependent

$$\text{(3)} \qquad \sum_j [a_j^* - a_j(-1)] = \Delta w$$

The coefficients of these variables consequently cannot be meaningfully interpreted nor empirically identified since no variable can change without the remaining variables also changing in any of an infinite number of ways to maintain the balance sheet identity. There is thus an infinite variety of values for the parameters of (2b) which are consistent with any observed behavior, and each individual parameter is meaningless in that any value is consistent with a particular unique behavioral specification.

A portfolio disequilibrium is fully described by any n of the $n+1$ variables in (2b) and all information relevant to an adjustment model is contained in n such variables. The assignment of specific values to n associated parameters will consequently uniquely describe the behavioral response of a particular asset holding in all possible portfolio situations. It does not require behavioral assumptions to select n variables that will describe the disequilibria, and the specific n variables that are used in the adjustment equations do not constrain behavior in any way. The choice of variables will of course define the interpretation of the individual associated coefficients, but any particular representation can be uniquely rewritten in terms of the coefficients of any other n selected variables.

---

[1] Similar arguments apply when some desired asset shares are assumed to depend upon only a subset of the explanatory variables.

Rather than simply choosing n variables to describe the disequilibria, a Pitfalls model could be formally constructed by substituting the identity (3) into the Clinton-Ladenson model (2b) in order to eliminate a redundant variable. Two of the more popular forms of the Pitfalls model will be illustrated here.

If Δw is eliminated from (2b), then the model can be written as

$$(3') \quad \Delta a_i = \sum_{j=1}^{n} (\epsilon_{ij} + f_i)[a_j^* - a_j(-1)] = \sum_{j=1}^{n} \theta_{ij}[a_j^* - a_j(-1)]$$

The adding up restrictions can be directly derived from

$$\Delta w = \sum_i \Delta a_i = \sum_i \sum_j \theta_{ij} \cdot [a_j^* - a_j(-1)] = \sum_i \theta_{ik}\Delta w + \sum_{j\neq k} \sum_i (\theta_{ij} - \theta_{ik}) \cdot [a_j^* - a_j(-1)]$$

This equation can hold for all possible discrepancies if and only if $\sum_i \theta_{ij} = 1, \ \forall j$. A coefficient θ_{ij} should be interpreted as the partial effect on holdings of the ith asset of a unit increase in desired holdings of the jth asset accompanied by an equal increase in Δw (since one discrepancy is increasing while the remaining discrepancies are held constant). The adding up restrictions state that such an increase in funds must be fully absorbed in the portfolio.

An alternative form of writing the Pitfalls model can be obtained by instead eliminating the kth discrepancy from (2b):

$$(3'') \quad \Delta a_i = \sum_{j\neq k} (\epsilon_{ij} - \epsilon_{ik})[a_j^* - a_j(-1)] + (f_i + \epsilon_{ik})\Delta w = \sum_{j\neq k} \phi_{ij}[a_j^* - a_j(-1)] + h_i \Delta w$$

The adding up restrictions follow from

$$\Delta w = \sum_i \Delta a_i = \sum_{j\neq k} \left(\sum_i \phi_{ij} \right) \cdot [a_j^* - a_j(-1)] + \left(\sum_i h_i \right) \Delta w$$

This can always hold if and only if

$$\sum_i \phi_{ij} = 0, \quad j \neq k \qquad \sum_i h_i = 1$$

Here the explicit coefficients ϕ_{ij} should be interpreted as the partial effect on holdings of the ith asset of desiring to hold one unit more of the jth asset and one unit less of the kth asset. Since Δw is held constant, the adding up restrictions require that any increases in asset holdings must be fully offset by decreases in other holdings. The h_i represents the partial effect of an increase in Δw which the sector desires to hold as the kth asset; the total portfolio will consequently expand by one unit.

The representations (3′) and (3″) are equivalent characterizations, and unique parameter values could be assigned in either case which would identically describe behavior in all portfolio situations. Specifically, the variables in (3′) could be rearranged as

$$\Delta a_i = \sum_{j=1}^{n} \theta_{ij}[a_j^* - a_j(-1)] = \sum_{j\neq k} (\theta_{ij} - \theta_{ik})[a_j^* - a_j(-1)] + \theta_{ik}\Delta w$$

Since this is the same form as (3″), it follows that

$$\phi_{ij} = \theta_{ij} - \theta_{ik}, \quad j \neq k \qquad h_i = \theta_{ik}$$

is necessary and sufficient to insure identical behavior no matter what the values of the explanatory variables. That is, one does not need to make behavioral assumptions in order to rearrange the explanatory variables. Note also that rearranging the variables and redefining the coefficients only rearranges the basic adding up restrictions.

II

After assuming that $\sum_i \epsilon_{ij}$ is constant over j, Clinton proves that

$$\sum_i (\epsilon_{ij} + f_i) = \sum_i \theta_{ij} = 1, \quad \forall j$$

is sufficient to insure adding up in adjustment model (2b). Completely ignoring the linear dependency in this model, he then offers the case

$$\epsilon_{ij} = \begin{cases} 1 - \sum_i f_i, & i = j \\ 0, & i \neq j \end{cases}$$

as a counterexample to Brainard and Tobin's statement that "If no cross-effects were allowed in the explicit equations of adjustment . . . , then the counterparts of all the own-adjustments specified would be loaded into the implicit adjustment equation for [the omitted assets]" (p. 106).

This statement in the Pitfalls paper follows a summary argument against own-adjustment equations such as (2a), the details of which I've given above, and refers to the consequences of omitting cross-effects from the Pitfalls model when written in the form (3″). Clinton's example applies to a different set of parameters and, in light of the arbitrariness of the parameter values in his equations, is neither surprising nor interesting.

What Clinton's example does illustrate is that one must be careful when discussing coefficients (such as "cross-effects") to specify which variables are being held constant. To be meaningful, Clinton's example has to be rewritten in terms of identifiable coefficients.

If, for instance, his example is written in the form of (3′), then

$$\Delta a_i = \left(1 - \sum_j f_j + f_i\right)[a_i^* - a_i(-1)] + \sum_{j \neq i} f_i [a_j^* - a_j(-1)]$$

The cross-effects (f_i) here specify that when one has a unit increase in available funds (w) accompanied by a unit increase in desired holdings of the jth asset, how much of these funds will go into other asset positions. Clinton has not eliminated these cross-effects, and indeed he could not eliminate them without assuming complete adjustment, $a_i = a_i^*$.

If Clinton's example is instead rewritten in the form (3″) that is used throughout the Pitfalls paper,

$$\Delta a_i = \left(1 - \sum_i f_i\right)[a_i^* - a_i(-1)] + f_i \Delta w, \qquad i \neq k$$

$$\Delta a_k = \sum_{j \neq k} - \left(1 - \sum_i f_i\right)[a_j^* - a_j(-1)] + \left(f_k + 1 - \sum_i f_i\right)\Delta w$$

The cross-effects now specify that when one wishes to switch 1 unit from asset k to asset j, how much will be put into asset i ($\neq j$). Instead of contradicting Brainard and Tobin, Clinton's example directly and completely confirms their statement. Precisely as Brainard and Tobin wrote, leaving cross-effects out of all but the kth equation necessitates cross-effects in this residual equation which fully offset the own-effects in the other equations. And cross-effects cannot be absent from all of the equations without eliminating the own-effects, $\Delta a_i = f_i \Delta w$.

In summary Brainard and Tobin are correct in their criticism of own-adjustment equations (2a) and in their discussion of the consequences of setting the cross-effects of model (3″) equal to zero. A similar argument has been made here against equating to zero the cross-effects of model (3′). None of this rules out someone arbitrarily calling some other parameters "cross-effects" and then assigning them a value of zero.

III

In contrast, Ladenson recognizes the linear dependency in the explanatory variables when he substitutes (1) into (2b) for estimation purposes

$$(4)\ \Delta a_i = \left[\sum_{j=1}^{n} \epsilon_{ij} b_{j0}\right] + w \sum_{h} \left[\sum_{j=1}^{n} \epsilon_{ij} b_{jh}\right] r_h w - \sum_{j=1}^{n} \epsilon_{ij} a_j(-1) + f_i \Delta w$$

where $w - \sum_{j=1}^{n} a_j(-1) = \Delta w$.

Although the coefficients can consequently not be uniquely estimated, Ladenson is nevertheless determined to assign values to all of them so that he may calculate estimates for all of the parameters in (1) and (2b). He therefore views the necessary elimination of a linearly redundant explanatory variable from his estimation equation (4) as an "expedient" parameter restriction with behavioral implications. He thus asserts that the elimination of w from (4) ". . . involves the assumption that assets adjust instantaneously to a change in wealth," whereas if Δw is eliminated ". . . we assume that such adjustment occurs with a distributed lag" (1973, p. 1006).

It is of course always possible to impose parameter constraints which alter the model; Ladenson (1971) demonstrates this by resorting to the expedient of setting $b_{j0} = 0$ for all j, even though this violated the consistency requirement that $\sum_j b_{j0} = 1$. However, it is certainly not necessary to impair the model in order to estimate it, and eliminating w, Δw, or $a_k(-1)$ from (4) in truth has no effect upon the model beyond a simple rearrangement of the identifiable parameters.

It was demonstrated above that there are a variety of equivalent and uniquely related ways in which the Pitfalls adjustment equations can be written; it will now be shown that there are similarly a variety of equivalent and uniquely related ways of writing the corresponding Pitfalls estimation equations. In both cases, the Ladenson alternative is to work with linearly dependent variables and unidentifiable parameters; and in both cases it is the elimination of a redundant variable rather than a behavioral assumption that is required to obtain identifiable parameters.

The substitution of (1) into form (3′) of the Pitfalls adjustment equations yields the estimation equation

$$(5a)\quad \Delta a_i = \left[\sum_{j} \theta_{ij} b_{j0}\right] w + \left[\sum_{h} \left[\sum_{j} \theta_{ij} b_{jh}\right] r_h\right] w - \sum_{j} \theta_{ij} a_j(-1)$$

This is indistinguishable from Ladenson's estimation equations (4) when Δw is eliminated. Behavioral restrictions were not needed to derive (3′) from Ladenson's (2b), and are consequently not needed to eliminate Δw from Ladenson's (4).

There are clearly an infinite number of equivalent ways in which (5a) could be rewritten. For example, substituting for $a_k(-1) = w - \sum_{j \neq k} a_j(-1) - \Delta w$ yields this arrangement of explanatory variables

$$(5b)\quad \Delta a_i = \left(\sum_{j} \theta_{ij} b_{j0} - \theta_{ik}\right) w + \sum_{h} \left(\sum_{j} \theta_{ij} b_{jh}\right) r_h w - \sum_{j \neq k} (\theta_{ij} - \theta_{ik}) a_j(-1) + \theta_{ik} \Delta w$$

This is Ladenson's (4) with $a_k(-1)$ eliminated, which could also be directly derived by substituting (1) into (3″) with the correspondence between the parameters again defined by

$$\phi_{ij} = \theta_{ij} - \theta_{ik}, \quad j \neq k \qquad h_i = \theta_{ik}$$

Since (3″) is equivalent to (3′), it follows that deleting $a_k(-1)$ from Ladenson's (4) is equivalent to deleting Δw.

Alternatively, substituting in either (5a) or (5b) for $w = \Delta w + \sum a_j(-1)$ yields yet another equivalent rearrangement of the variables which this time is indistinguishable from Ladenson's (4) with w eliminated.

All of these cases (and all alternative linear rearrangements of the variables and parameters) are just different ways of writing the same model. If one is careful, the form in which the model is written has absolutely no behavioral implications. In particular, irrespective of whether w, $a_k(-1)$, or Δw is eliminated from (4) and no matter how the resultant linearly independent variables are juggled, the final equations will be equivalent to form (5a), and the coefficients of the rearranged variables will be uniquely related to the parameters of (5a). Indeed, it would be very strange if one were only happy with the model when the variables were arranged in a particular way; again we have the Pitfalls moral that explicit representations are no more important than implicit ones.

IV

Single equation ordinary least squares (*OLS*) parameter estimates will generally obey adding up restrictions if the regressors are the same in each equation and if some linear combination of the regressors is always equal to the sum of the regressands.[2] This can be seen directly by writing an estimation equation as

$$\begin{array}{ccccccc} a_i & = & X & \cdot & \beta_i & + & \epsilon_i \\ T\text{x}1 & & T\text{x}p & & p\text{x}1 & & T\text{x}1 \end{array}$$

where we have (possibly) rearranged X so that the first regressor is

$$x_1 = \sum_{i=1}^{n} a_i$$

The sum across equations of the *OLS* parameter estimates are the elements of the column vector

$$\sum_{i=1}^{n} \hat{\beta}_i = \sum_i (X'X)^{-1}X'a_i = (X'X)^{-1}X' \sum_i a_i$$

which is what would result from a regression of $\sum a_i$ on X. The squared residuals for that regression are minimized (are equal to zero) when the coefficient of x_1 is one and all other coefficients are zero.

For *OLS* estimation of the model considered here, the Gauss-Markov Theorem implies that linearly rearranging the explanatory variables in (5a) will not affect the forecasts nor the implicit estimates of any estimatable parameters. In particular, the implicit estimates of the coefficients of (5a) do not depend upon whether w, Δw, or $a_k(-1)$ is deleted from Ladenson's (4). Thus, Ladenson's alleged behavioral constraints are just awkward ways of eliminating a redundant variable and differ only in providing different arbitrary estimates of uninteresting parameters.

In considering *OLS* estimates of the identified structural parameters, it consequently suffices to consider the single case of using the estimation equations (5a) to estimate the parameters of the Pitfalls model (1) and (3'). Rewriting (5a) in obvious matrix notation,

$$a = \theta \cdot B_0 \cdot w + \theta \cdot B \cdot rw + (I_n - \theta)a(-1) + \epsilon$$

(where a is nx1, θ is nxn, B_0 is nx1, B is nxq, and r is qx1). The column elements of the estimates $\widehat{\theta B_0}$ will sum to one and the elements of $\widehat{\theta B}$ and $(\widehat{I-\theta})$ will sum to zero. The column elements of $\hat{\theta} = (\widehat{I-\theta}) + I$ will thus sum to one as desired. The elements of $\hat{B}_0$ and $\hat{B}$ will also sum correctly, since it can be easily shown that the column sums for any matrix $Z = \theta^{-1}(\theta Z)$ will be the same as the column sums of θZ if the columns of θ sum to one.

It is worth noting that hypothesis tests of any coefficients or combinations of coefficients are also invariant to the way in which the equations are written for estimation purposes. However, the usual computer output provides t-statistics for testing the null hypotheses that the explicitly estimated parameters are equal to zero. If one is in the (dubious) habit of imposing zero constraints whenever these are unrejected by the data, then the form in which the equations are written may affect the final estimates if one allows this arbitrary choice to define the subset of possible hypotheses that will be tested. This is of course not a limitation of the model but rather a well-known drawback to mechanically using hypothesis tests for decision-making purposes.

Turning to the question of more complex estimation methods, the automatic imposition of adding up constraints obviously extends to a variety of single equation methods, such as the use of instrumental variables and semi-first differences if and only if the instruments used for each variable and the autoregressive parameters do not vary across equations. It will also remain true that the properly interpreted estimates are independent of the particular linearly dependent variable omitted from (4) and of any subsequent rearrangement of the variables.[8]

---

[2] Ladenson incorrectly states that it is necessary and sufficient that all lagged endogenous variables appear in all equations.

For the class of feasible Aitken generalized least squares system estimators, we will generally have to delete one equation (since the disturbances must sum to zero) and calculate the parameter estimates for this residual equation from the adding up constraints. Fortunately Alan Powell has shown that the estimated coefficients do not depend upon the equation chosen for deletion. In addition, it is well known that such system estimates will be identical to the single equation estimates on which they are based if the final step in the single equation method is the *OLS* regression of equations in which the right-hand side variables do not vary across equations.

The likelihood that the effective dimensionality of the data will be much less than the number of parameters is a persuasive argument for the imposition of a priori parameter restrictions. If one does specify the coefficients associated with a particular variable in every equation, then the single equation methods discussed above will continue to automatically impose the adding up constraints, and the estimated parameters will be identical to the ones that would result from an Aitken procedure. If the parameters are not specified in every equation, then one can enforce the adding up constraints with a stacking procedure or with a Lagrangian modification of the Aitken method (see the discussion by Gordon Sparks for example). If one has useful information about the reliability of the parameter restrictions then a Bayesian or mixed estimation method might be of interest (see Smith and Brainard for example). Unfortunately, all of the non-single equation methods typically involve formidable computational difficulties and the possibility of serious rounding errors.

REFERENCES

W. Brainard and J. Tobin, "Pitfalls in Financial Model Building," *Amer. Econ Rev. Proc.*, May 1968, *58*, 99–122.

K. Clinton, "Pitfalls in Financial Model Building: Comment," *Amer. Econ. Rev.*, Dec. 1973, *63*, 1003–04.

M. Ladenson, "Pitfalls in Financial Model Building: Some Extensions," *Amer. Econ. Rev.*, Mar. 1971, *61*, 179–86.

———, "Pitfalls in Financial Model Building: Reply and Some Further Extensions," *Amer. Econ. Rev.*, Dec. 1973, *63*, 1005–08.

A. Powell, "Aitken Estimators as a Tool in Allocating Predetermined Aggregates," *J. Amer. Statist. Assn.*, Sept. 1969, *64*, 913–22.

G. Smith and W. Brainard, "The Value of A Priori Information in Estimating a Financial Model," paper presented at Dec. 1973 meetings of the Econometric Society.

G. Sparks, "An Economic Approach to the Analysis of Portfolio Decisions: The Demand for Financial Assets in the United States," mimeo.

---

[8] With instrumental variables, this conclusion requires that all predetermined variables in the equation be included in the list of instruments. We can then view all of the explanatory variables as being regressed upon a common list of regressors. By the earlier adding up proof, forecasts of linear combinations of the explanatory variables will then equal linear combinations of the forecasts of the individual explanatory variables.

Buyer's Prices, Seller's Prices, and Price Flexibility: Comment

By DOUGLAS R. BOHI AND GERALD W. SCULLY*

Recently, the controversy over the administered-price thesis has been renewed with the publication in a book by George Stigler and James Kindahl of a new set of industrial price indexes developed under the auspices of the National Bureau of Economic Research. The National Bureau (*NB*) price indexes are based upon contract prices paid by the buyer. In contrast, the Bureau of Labor Statistics (*BLS*) price indexes employ price quotations or price lists obtained from the seller. Stigler and Kindahl conclude that the *NB* price indexes reveal a significantly different pattern of price behavior from the *BLS* indexes, particularly in the behavior of short-run movements, where the *NB* index changes more often and more smoothly than the *BLS* index.

Furthermore, Stigler and Kindahl employ these buyer-based price indexes to test their version of the administered-price thesis "... that contractions in business lead to no systematic reduction of industrial prices, and ... expansions in business may only tardily lead to price increases" (pp. 7–8). They find that the new commodity price indexes move in response to general business activity and that there is "no evidence here to suggest that price rigidity or 'administration' is a significant phenomena" (p. 9).

In contrast, Gardiner Means recently reexamined the *NB* price data and concluded "... that the new data overwhelmingly support the administered-price thesis; ... and that the new data do not support the report's conclusion that the prices in its sample show a tendency to move with the business cycle" (p. 292).

There are two principal testable hypotheses which emerge from this important controversy. First, there is the question whether the *NB* and *BLS* price indexes are actually different, and, secondly, there is the issue of whether industrial prices respond to market changes over the business cycle or whether they are in fact administratively determined. Both issues are concerned with the nature of the variance of industrial prices. Both the Stigler and Kindahl and the Means studies have analyzed the variance of industrial prices solely within the time domain.

To establish the dissimilarity of the *NB* and *BLS* price series, Stigler and Kindahl used the simple correlation coefficient of unlagged first differences in the logarithms of the indexes over frequency intervals ranging from one month to one year. To test the administered-price thesis, both Stigler and Kindahl and Means examined the average price change occurring within the two reference cycles in the period of coverage. Both of these tests are weak. Furthermore, according to Stigler-Kindahl, buyer price changes are expected to lag seller price changes and price changes are expected to lag behind changes in industrial production; hence, both tests are likely to be biased. In particular, the bias will be toward accepting the dissimilarity of the two price indexes and towards accepting the administered-price thesis.

We can suggest alternative time domain tests of the hypotheses. First, the relationship between the *NB* and *BLS* indexes might be analyzed using the following regression:

$$\Delta NB_{ij} = \beta_{ij}\Delta BLS_{ij} \tag{1}$$

$$i = 1, n;\ j = 1, m$$

where i denotes the industry and j the time lag in months. The regression coefficient would be a measure of the average amplitude difference between the two series. Appropriate tests would be whether $\beta_{ij}>0$ and $\beta_{ij}>1$.

---

* Associate professor of economics, Southern Illinois University and professor of economics, Southern Methodist University, respectively. Scully's research was partially financed by PHS Research Grant No. 1-R01-HD05989-01 from the National Institute of Health and Institutional Grant No. 31-46-70-06 from the Manpower Administration, U.S. Department of Labor.

Second, to provide some insight into the administered-price thesis, a more powerful alternative to a simple examination of direction and magnitude of price changes within reference cycles would be the following relation between changes in industrial prices (Δp) and changes in the index of industrial production (ΔIP)

$$\Delta p_{kij} = \alpha_{kij} \Delta IP_{ij} \tag{2}$$

where k is the NB or BLS index, i the industry, and j the time lag. We would test whether $\alpha_{kij}>0$ and $\alpha NB_{ij}=\alpha BLS_{ij}$. Such a test potentially would establish whether the contour of price changes in the ith industry over the j-th time interval is correlated positively with the industrial production index IP and whether the amplitude differences in price behavior are significantly different between the NB and BLS indexes. However, while both time domain tests (1) and (2) potentially offer more powerful insights into the controversy, they suffer the same inherent problems of the weaker tests already undertaken by Stigler and Kindahl and Means. Unless the lags are appropriately specified, the tests will be biased, and there is no a priori basis for selecting one time lag over another.

In this paper, the important hypotheses in the debate are examined in the frequency domain, using spectral and cross-spectral techniques.[1] The use of a frequency domain approach to both the hypothesis of dissimilarity of the NB and BLS price indexes and the administered-price thesis is justified on the grounds that 1) the two issues essentially involve the relative frequency characteristics of different time-series and therefore the relevant hypotheses are conveniently formulated and tested in the frequency domain; 2) the frequency domain statistics to be used are not dependent upon a specification of the time lag (although some statistics, such as the phase, would require such a specification) and therefore avoid the time domain bias already discussed; and 3) spectral methods provide a convenient and efficient method for separating the variance of a time-series into cycles of different lengths and thus permit a comparison of variances over short, intermediate, and long-run periods.

I. *BLS-NB* Price Comparison

The *BLS* industrial price indexes are constructed from seller price lists, while the *NB* indexes are based on buyer prices. Stigler and Kindahl assert that the *BLS* procedure is deficient in several respects. The *BLS* indexes fail to capture price movements which result from discounts, changes in the terms of trade, and special concessions.[2] In addition, the *BLS* indexes are based upon a relatively small number of price reports (an average of one to three), while the number of reporters is a major determinant of the number of reported price changes.[3] The *BLS* procedure suggests a unique market price in which the buyer may shift from one supplier to another in response to better terms. The seller's price is for a precisely defined commodity, regardless of whether or not any is sold, while the buyer's prices reflect a wider range of commodities purchased continuously over time. These differences, among others, lead Stigler and Kindahl to conclude that "the seller's price quotations . . . will not display the multiplicity of prices and the frequency of revision that the transactions prices actually display" (p. 39).

To support their claim, Stigler and Kindahl examined 64 *BLS* and *NB* price indexes over the ten-year period from 1957 to 1966. The indexes were compared with respect to trend, cyclical fluctuations, and short-run movements. They find that slightly over half of the *BLS* indexes had steeper trends than the corresponding *NB* indexes, while the remainder were split between equal trends and steeper trends in the *NB* indexes, pp. 56–59. Comparing price changes during the reference cycle contractions (July 1957–Apr. 1958 and Jan 1960–Jan. 1961), they discover that "There are two *NB* price movements in the expected (downward) direction for each perverse change; in the *BLS* indexes the expected and perverse movements were virtually equal" (p. 61). When compared during two reference cycle expansions (Apr. 1958–June 1959 and Jan. 1961–Mar. 1962), "The *BLS* and *NB* price indexes . . . are essentially identical in their behavior" (p. 63). Finally, with regard to short-run movements, Stigler and Kindahl report that ". . . the *BLS* indexes usually change less frequently than the *NB* indexes. . . . But the movements of the *BLS* indexes are large and jerky" (p. 66). Correlating the first differences in the logarithms of the indexes, they find ". . . no close agreement in the monthly movements of the two indexes . . . the mean correlation coefficient is only .32 . . ." (p. 67). The mean correlations increase to .45 for quarterly changes, .54 for semiannual changes, and .59 for annual changes Furthermore, Stigler and Kindahl conclude that ". . . the quoted prices lag behind the changes in transactions prices by an unknown but appreciable time interval. The *NB* and *BLS* indexes are not very closely related in timing . . ." (p. 70).

---

[1] See Clive Granger and Michio Hatanaka or G. S. Fishman for a detailed discussion of spectral methods. For a more heuristic description of the technique, see Scully.

[2] Stigler and Kindahl also note that buyers were reluctant to disclose purchase prices for which an ". . . extraordinary favorable deal had been consummated" (p. 27).

[3] In contrast, the *NB* indexes used a total of 1,300 price reporters for 70 commodities, or an average of 17. Stigler and Kindahl add: "This heavier coverage reflects our distrust of the belief that there is only one price in a market at a given time" (p. 5).

The Stigler-Kindahl statistical comparisons of the *BLS* and *NB* price indexes suggest widespread differences in the behavior and character of these price series. However, neither Stigler and Kindahl nor Means have sought to test directly the hypothesis that the two series are obtained from essentially different stochastic processes. We propose to perform such a test using spectral techniques. Spectral density functions are decompositions of variance into frequency bands. The sum of the spectral density function is in fact the total variance in the time-series. We employ spectral methods because we are interested in the lag-adjusted behavior of these price indexes in several frequency bands, i.e., low, moderate, and high frequencies. In the spectral comparisons of the two indexes that follow, the spectral statistics were obtained using a Tukey-Hannan window with a maximal lag of 30 months and the sample data provided in the Appendix of the Stigler-Kindahl book. In the first test of similarity, spectral density functions were estimated for each of the 64 *BLS* and *NB* commodity indexes and then the power at each frequency of each *BLS* index was divided by the corresponding power of each *NB* index. Let $Q_i(w)$ represent this ratio of power at frequency w for commodity i and let $P_i(w)$ represent the ratio of the "true" power densities at frequency w for commodity i. The following $(1-\alpha)$ confidence interval holds:

$$P\left(F_{1-\alpha/2} \leq \frac{Q_i(w)}{P_i(w)} \leq F_{\alpha/2}\right) = 1 - \alpha$$

where $F_{\alpha/2}$ and $F_{1-\alpha/2}$ are the values from the F-distribution for α level of significance and (8, 8) degrees of freedom.[4] The null hypothesis to be tested is $P_i(w)=1$; that is, there is no significant difference between the spectra of the *BLS* and *NB* indexes at each frequency or that the two indexes simply reflect the same stochastic processes. Assuming the null hypothesis to be true, we should expect

$$P(F_{1-\alpha/2} \leq Q_i(w) \leq F_{\alpha/2}) = 1 - \alpha$$

However, the $(1-\alpha)$ confidence interval is somewhat wider with respect to all frequencies.[5] In particular, if we set the error rate at $.05/(m+1)$ where m is the number of lags (that is, a 5 percent chance of error at each frequency), the joint probability of committing any error over all $(m+1)$ frequencies will be no greater than .05.[6] For

---

[4] The number of degrees of freedom is given by $2n/m$ where n is the number of sample observations (120) and m the maximal lag (30).

[5] See Thomas Naylor, Kenneth Wertz, and Thomas Wonnacott for a discussion of this test. We thank Anthony Lawrence for calling our attention to this.

[6] The joint probability of an error over $(m+1)$ outcomes will not exceed the sum of the probabilities of

TABLE 1—CLASSIFICATION OF COMMODITY INDEXES BY *NB* INDUSTRY CODE ACCORDING TO DISSIMILARITY OR SIMILARITY OF *NB* AND *BLS* SPECTRA

Cases Where *NB* and *BLS* Spectra Were Not Significantly Different (46 cases)					Cases Where *NB* and *BLS* Spectra Were Significantly Different (18 cases)	
C-1[b]	C-10[a]	C-22	C-41	C-51	C-16[a]	C-36
C-2	C-11[a]	C-29	C-42	C-52	C-20	C-37
C-3	C-12[a]	C-30	C-43	C-54	C-21	C-39
C-4	C-13[a]	C-31	C-44	C-57	C-23[a]	C-40[b]
C-5	C-14[a]	C-32[b]	C-45	C-58	C-24	C-50
C-6	C-15[a]	C-33	C-46	C-59	C-25	C-53[b]
C-7	C-17[a]	C-34	C-47	C-60	C-26	C-55
C-8[b]	C-18[a]	C-35	C-48[b]	C-61	C-27[b]	C-56
C-9	C-19[a]	C-38[b]	C-49[b]	C-62	C-28[b]	C-63[a]
				C-64[a]		

[a] Market dominated, according to Means, p. 306.

[b] Nonconforming to the business cycle, according to Means, p. 305.

$\alpha = .05$ and $m = 30$, the area in each tail of the F-distribution becomes .025/31, or .0008. Thus, the simultaneous confidence interval becomes

$$P(F_{.9992} \leq Q_i(w) \leq F_{.0008}) \leq .95$$

Of the total of 64 commodity indexes available for analysis, 46 had $Q_i(w)$ ratios that consistently fell within the 95 percent confidence bands. We conclude that the variance patterns of these 46 commodities are not significantly different whether measured by the *BLS* method or the *NB* method. This holds true for both long-run and short-run cyclical movements. Table 1 lists the commodity indexes that on this basis are judged not significantly different along with the remaining 18 commodity indexes where the $Q_i(w)$ ratios exceed the 95 percent confidence bands. Commodity groupings which exhibit the greatest similarity in the two indexes are Metals and Products, Paper Products, and Nonmetallic Mineral Products. The greatest differences in the spectra between the *BLS* and *ND* series are among Fuel Products, Rubber Products, and Chemical Products. Among the latter 18 commodities, it is of interest to note that 12 commodities or two-thirds showed the variance of the *BLS* index to be substantially greater than that of the *NB* index. Thus, only in 6 out of 64 cases was the variance of the *NB* index greater than that of the *BLS* index. Furthermore, this greater variance was not restricted to any particular frequency range but was verified in both high and low frequencies.

The coherence between the *BLS* and *NB* indexes shows a pattern similar to the correlations reported by Stigler and Kindahl. The coherences generally were larger in the lower than in the higher frequencies. As expected, the coherences were significantly greater for those commodities where the spectra were similar. Among this group, average coherences across all frequencies were in excess of .5. The average coherences among the remaining commodities generally were below .5.

In spite of the numerous similarities among the *BLS* and *NB* price indexes, it is important to consider whether the similarities tend to occur among commodities where prices are relatively constant or among commodities where prices are highly flexible. Accordingly, indexes of relative price variance were constructed for each of the 64 commodities by dividing the spectral density functions of each commodity by the spectral density functions of the aggregate industrial

---

each outcome. But the sum of $(m+1)$ probabilities, all with the same $.05/(m+1)$ probability, is simply .05. See Naylor et al.

price index. Four such indexes were obtained for each commodity, one for total variance (encompassing all frequencies), one for the variance attributed to the low frequencies (cycles having a period greater than 12 months), one for the variance attributed to the intermediate frequencies (cycles with a period between 3 and 12 months), and one for the variance attributed to the high frequencies (cycles with a period less than 3 months). The commodity indexes were ranked from highest to lowest on the basis of the relative variance. The rankings obtained on the basis of this procedure for the *BLS* price indexes were compared with those obtained for the *NB* price indexes.

The rank correlations between the variance rankings of the *BLS* and *NB* indexes are presented in Table 2. The rank correlations are all highly significant, although it is to be noted that the correlations decrease with the higher frequencies. The total frequency rankings were almost identical to the low frequency rankings, reflecting the common phenomenon among economic time-series that the bulk of the variance is attributable to the longer cycles.

TABLE 2—RANK CORRELATIONS BETWEEN THE VARIANCE RANKINGS OF THE *BLS* AND AND *NB* PRICE INDEX[a]

Rank Correlation		t-statistics
Total Frequencies	.754	8.962
Low Frequencies	.763	9.216
Intermediate Frequencies	.673	7.105
High Frequencies	.631	6.349

[a] The actual variance rankings for the 64 commodities over the various frequency intervals are available from the authors upon request.

We have also examined the variance ranking of each commodity index. Considering first the low frequencies of the top 20 ranked indexes (that is, the frequencies accounting for the bulk of the variance of the price indexes exhibiting the greatest variance), it is noted that nearly all of the commodities are in the group where the *NB* and *BLS* spectra are similar and that, with few exceptions, the top 20 indexes are the same commodities whether measured by the *BLS* or *NB* method. The notable exceptions include Insulated Copper Wire (C-16), which is ranked in the top ten by both measures but nevertheless had dissimilar spectra; Cardiac Glycosides (C-56), which ranks first using the *BLS*-method but 28th with the *NB* method; and Passenger Tires (C-24), which ranks 16th among the *BLS* indexes and 43rd among the *NB* indexes. All exceptions involved commodity indexes with dissimilar power spectra.

The same pattern holds for the high frequencies of the top 20 ranked indexes, though with less consistency. In fact, the same commodities generally appear among the top 20 in both the low and high frequencies. Again, the exceptions to the pattern develop among those commodities with dissimilar power spectrums.

The pattern is reversed among the 20 commodities with the smallest variances. In both the low and high frequencies, nearly half of these indexes have dissimilar power spectra. Furthermore, about half of the lowest 20 ranked *BLS* indexes are not the same commodities that appear among the lowest 20 ranked *NB* indexes. The most notable example here is Plywood (C-63), which is ranked seventh among the *BLS* indexes in the high frequencies, but sixtieth among the *NB* indexes.

These results suggest that there is considerable agreement and similarity between the *BLS* and *NB* methods of calculating price indexes when they are being used to measure price changes that are relatively large. The major differences between the *BLS* and *NB* indexes occur among those commodities where there is relatively little price variability.

II. Price Flexibility

According to Means, the "... administered-price thesis held that in business recessions administered-prices showed a tendency not to fall as much as market prices while the recession fall in demand worked itself out. ... Similarly, ... they tended not to rise as much in recovery ..." (p. 292). Thus, administered prices are held to be relatively inflexible over the course of the business cycle and perhaps, may move countercyclically. Stigler and Kindahl, however, interpret the thesis strictly in terms of the conformity of price movements to that of the general business cycle: "The main thrust of the doctrine of administered prices is that contractions in business lead to no systematic reduction of industrial prices, and much more equivocally, expansions in business may only tardily lead to price increases" (p. 7).

It is worth emphasizing that the administered-price thesis does not imply that industrial prices are inflexible with regard to factors other than transitive shifts in aggregate demand. That is to say, structural shifts in demand among industries or changes in cost functions in specific industries produce price changes generally unrelated to the business cycle and, as such, are not a matter of concern to the administered-price thesis. Changes in aggregate demand, on the other hand, are transitory phenomena in which different industries may respond variously through an inventory policy or price policy. Administered-price industries would tend to respond by varying output or inventories; other industries by altering prices.[7]

After examining the *NB* price data, Means disputes the Stigler and Kindahl conclusion that industrial prices tend to move in response to the business cycle. Means regards 13 of the 64 commodity indexes as market dominated. He shows that 30 of the remaining 51 failed to move in conformity with changes in business activity in three or four times out of a possible four opportunities (i.e., the contractions and recoveries of 1958–59 and 1961–62, (pp. 304–04)). However, of these 30 only 10 were completely perverse in all four opportunities. Moreover, Means argues that the prices of the 13 market-dominated commodities tend to behave "quite differently" relative to the administration-dominated prices. ". . . thus confirming the administered-price thesis" (pp. 303–04).

It is of interest to note that of the 13 commodities Means designated as market dominated, 10 are in the group of Table 1 where the *BLS* and *NB* spectra are not significantly different. Furthermore, of the 10 commodities that were completely perverse with respect to the business cycle, 6 were among the 46 commodities with similar spectra and 4 were in the group with dissimilar spectrums.

The average rank of the two groups is significantly different as well. The market dominated commodities tend to fall among the top 20 variances, with an average ranking of 19–20 in both the high and low frequencies. The nonconforming commodities, on the other hand, tend to have much lower variances, with an overall average ranking between 39 and 43. Thus, the results tend to confirm Means' version of the administered-price thesis which emphasizes relative price inflexibility.

The Stigler-Kindahl version of the administered-price thesis emphasizes the conformity of price movements to the business cycle. To test the Stigler-Kindahl version, we estimated cross-spectral density functions of the *BLS* and *NB* price indexes using the index of industrial production as the base series. The industrial production index is widely regarded as the single best indicator of the business cycle. The main statistic of interest here is the coherence square, which is a measure of the degree of association between the two series calculated at each frequency. The coherence relationship between each commodity price index and the industrial production index is described as either very low, low, moderate, or high depending on the magnitude and consistency of the coherences over all 30 frequencies (see the footnote to Table 3 for definitions). The sign of the cospectrum is reported to indicate whether a significant coherence relationship is inverse (negative) or direct (positive). Low and very low coherences imply that the two series are not sufficiently related to consider the sign and are designated as not applicable (*n.a.*). The last two columns of Table 3 indicate whether a significant coherence relationship is in-phase (*IN*) or out-of-phase (*OUT*). Two series are considered in-phase if the quadrature spectrum is consistently less than the cospectrum.

---

[7] We have investigated the behavior of output amplitudes in specific administered-price industries. These results are not presented here for two important reasons. First, the *BLS-NB* industrial price classifications differ substantially from the Federal Reserve Board industrial production classification, raising doubts about comparability. Second, where price adjustments are avoided in favor of output adjustments, inventories ought to be the primary focus of attention. Inventory data by volume are even more scarce than output data. Furthermore, without knowledge of the costs of holding and adjusting inventory stocks, interindustry comparisons of inventory adjustments over the business cycle would be inappropriate.

A price series which is completely perverse with respect to the index of industrial production would be one in which the coherence is moderate or high, the sign of the cospectrum is negative, and the two series are in-phase. A price series which consistently conforms to the cyclical pattern of industrial production differs from that of the

TABLE 3—COHERENCE, SIGN OF COSPECTRUM, AND PHASE OF *NB* AND *BLS* PRICE INDEXES COMPARED TO THE INDUSTRIAL PRODUCTION INDEXES FOR CASES WHERE COHERENCE WAS MODERATE OR HIGH[a]

	Coherence Description[b]		Sign of Cospectrum[c]		In/Out Phase[d]	
NB Index Number	*BLS*	*NB*	*BLS*	*NB*	*BLS*	*NB*
C-1[e]	*L*	*M*	*n.a.*	–	*n.a.*	*OUT*
C-5	*L*	*M*	*n.a.*	–	*n.a.*	*IN*
C-7	*M*	L	–	*n.a.*	*IN*	*n.a.*
C-12[f]	*L*	*H*	*n.a.*	+	*n.a.*	*IN*
C-13[f]	*M*	*H*	+	+	*IN*	*IN*
C-14[f]	*H*	*H*	+	+	*IN*	*IN*
C-15[f]	*H*	*H*	+	+	*IN*	*IN*
C-16[f]	*H*	*M*	+	–	*IN*	*OUT*
C-17[f]	*M*	*M*	+	+	*IN*	*IN*
C-18[f]	*M*	*M*	+	+	*IN*	*IN*
C-19[f]	*H*	*H*	+	+	*IN*	*IN*
C-25	*VL*	*H*	*n.a.*	–	*n.a.*	*IN*
C-26	*VL*	*M*	*n.a.*	–	*n.a.*	*IN*
C-33	*VL*	*M*	*n.a.*	+	*n.a.*	*IN*
C-35	*M*	*M*	+	+	*IN*	*IN*
C-37	*M*	*M*	–	–	*IN*	*IN*
C-38[e]	*VL*	*H*	*n.a.*	–	*n.a.*	*IN*
C-40[e]	*VL*	*M*	*n.a.*	–	*n.a.*	*IN*
C-41	*L*	*M*	*n.a.*	–	*n.a.*	*OUT*
C-42	*VL*	*M*	*n.a.*	–	*n.a.*	*IN*
C-44	*M*	*L*	+	*n.a.*	*IN*	*n.a.*
C-55	*L*	*M*	*n.a.*	–	*n.a.*	*IN*
C-58	*L*	*M*	*n.a.*	–	*n.a.*	*IN*
C-60	*L*	*H*	*n.a.*	+	*n.a.*	*IN*
C-62	*H*	*L*	–	*n.a.*	*OUT*	*n.a.*
C-64[f]	*VL*	*H*	*n.a.*	+	*n.a.*	*IN*

[a] Only those cases where the *NB* or *BLS* price indexes were moderately or highly coherent with the Industrial Production Index are reported. In all of the unreported cases the coherences were low or very low.

[b] The coherence descriptions are defined as follows: Very Low (*VL*): most coherences are below .25; Low (*L*): most coherences below .4 with a few above .4; Moderate (*M*): most coherences between .3 and .6; High (*H*): some coherences above .8, most above .5. (See Clive Granger and Oskar Morgenstern, p. 20.)

[c] The signs indicate that the coherence is negative or positive, while *n.a.* means that the sign is not relevant because the coherence is too low.

[d] The coherences are labelled in-phase or out-of-phase depending on the magnitude of the quadrature spectrum relative to the cospectrum. Again, *n.a.* indicates that the phase relationship is not relevant because of the lack of coherence.

[e] See fn. b, Table 1.

[f] See fn. a, Table 1.

above only with respect to the sign of the cospectrum. When the coherence is low, prices may be positively or negatively related to a given upswing or downswing in industrial production (such as the two cycles considered by Means and Stigler and Kindahl), but the overall coherence may be too low to reveal a consistent and significant cyclical relationship. This may be considered a shortcoming of an analysis which considers the cross-spectral results across all frequency bands rather than a selected few. However, two cycles hardly provide sufficient degrees of freedom on which to base statistical conclusions. Furthermore, the spectral density function of industrial production reveals a rather smooth downward sloping curve typical of economic time-series, where the variance attributed to cycles with a period of around two years is not sufficiently large to warrant a concentration of emphasis.[8] In fact, the lack of important peaks in the spectral density of industrial production suggests the necessity for emphasizing the cross-spectral results over all frequency bands (as in Table 3) rather than concentrating on individual frequency bands.

The most notable feature of the results in Table 3 is the large proportion of price series that have no significant coherence with the index of industrial production. Measured by the *BLS* method, 52 out of 64 indexes have low or very low coherences, while 41 of the *NB* indexes are so described. Thus, the bulk of these so-called administered-price series has no apparent cyclical relationship with industrial production. It is interesting to note, moreover, that the *BLS* indexes present this picture more definitively than the *NB* indexes. The lack of coherence between prices and output in general is consistent with Means' view of the administered-price thesis. However, the results do suggest the danger of basing such a conclusion on the lack of conformity of prices with just two swings in the business cycle.

Of the indexes which are significantly coherent with industrial production, there are some rather interesting differences in the *NB* and *BLS* indexes. These differences may partially account for the opposite conclusions reached by Means and by Stigler-Kindahl. The 12 *BLS* indexes that exhibited moderate or high coherences include 9 with positive cospectra and only 3 with negative cospectra. Thus, of those prices coherent with industrial production, the bulk move in conformity with output rather than countercyclically. However, of the 23 *NB* indexes with significant coherences, only 12 had positive cospectra. It is somewhat ironic that the *BLS* indexes support the Stigler-Kindahl position while their own price indexes provide no useful information one way or the other on this issue.[9]

Another feature of these results is that of the 9 *BLS* and 12 *NB* indexes that show conformity with the index of industrial production, 7 *BLS* and 8 *NB* indexes are among the group of 13 commodities that were labeled by Means as market dominated. The remainder had insignificant coherences. On the other hand, of the 3 *BLS* and 12 *NB* indexes with negative cospectra, none of the *BLS* and only 3 of the *NB* indexes were in the group of 10 commodities Means cites as completely perverse with respect to the business cycle. Otherwise, these commodities showed insignificant coherences.

Finally, the *BLS* and *NB* indexes generated about the same pattern of coherences with industrial production. Thirty-seven commodities had low or very low coherences with both indexes, while 9 commodities had moderage or high coherences with both indexes. It was expected that some differences in coherences would arise in cases where the power spectrum was significantly different between the *BLS* and *NB* indexes (as described in Table 1). Strangely enough, however, only 5 of the 18 commodities with different coherences had significantly different power spectra, which is no different than one would expect from a random drawing.

---

[8] See Scully for a discussion of the spectral shape of the industrial production index.

[9] Note, however, that 3 of the 13 *NB* indexes with negative cospectra are out-of-phase with output and may therefore appear in the time domain to move positively with output.

III. Conclusion

On the basis of the statistical results obtained by employing spectral and cross-spectral techniques to the *NB* and *BLS* industrial price series, we conclude that 1) for the most part, these two series reflect similar stochastic processes, and 2) administered prices do appear to be a feature of the *U.S.* economy. However, these conclusions require some qualification in view of the lack of universal evidence. Our analysis shows that nearly three-fourths of the Stigler-Kindahl price indexes have spectra similar to those of the *BLS* price series. Where the two indexes do differ, the commodity prices involved generally show relatively little variance anyway. Researchers who employ industrial price indexes can be indifferent between the *NB* and *BLS* series except for the cases we have noted.[10] It would not seem very profitable to follow the Stigler and Kindahl recommendation and proceed to gather information on buyers' prices universally. The Stigler-Kindahl conclusion of a general difference in the behavior of their price series in comparison to the *BLS* series would not appear warranted.

With regard to whether administered prices are a feature of the *U.S.* economy, we find little evidence to support the Stigler-Kindahl contention of conformity between the price indexes of these industries and the business cycle. Among the *BLS* series, 80 percent showed no conformity with the business cycle. Among the *NB* prices, the Stigler-Kindahl version of the administered-price thesis did somewhat better, but still 41 of 64 series had low or very low coherence with the industrial price index. Furthermore, the proportion showing nonconformity with the business cycle rises to three-quarters when the market-dominated commodities are removed from the list. Where there was conformity in the *NB* price series with the business cycle, nearly half the cases show a perverse correlation. Actually, it is the *BLS* indexes which more strongly support the Stigler-Kindahl argument. Here, three-quarters of the significantly coherent cases indicated cyclical conformity of price and output. But, it is important to note that 6 of 9 positive cospectra of the *BLS* series and 7 of the 12 positive cospectra of the *NB* series actually were market-dominated commodities. Hence, evidence of cyclical conformity of prices and output in administered-price industries appears very thin indeed.

---

[10] George Hay used the *NB* price indexes to reestimate certain equations in his model. The goodness of fit statistic for the price equation improved only in the Paper industry.

REFERENCES

G. S. Fishman, *Spectral Methods in Econometrics*, Cambridge, Mass. 1969.

C. W. J. Granger and M. Hatanaka, *Spectral Analysis of Economic Time Series*, Princeton 1957.

——— and O. Morgenstern, "Spectral Analysis of New York Stock Market Prices," *Kyklos*, 1963, *16*, 1–24.

G. A. Hay, "A Note on the Stigler-Kindahl Study of Industrial Prices," *Amer. Econ. Rev.*, Sept. 1972, *62*, 656–58.

G. Means, "The Administered-Price Thesis Reconfirmed," *Amer. Econ. Rev.*, June 1972, *62*, 292–306.

T. H. Naylor, K. Wertz, and T. H. Wonnacott, "Spectral Analysis of Data Generated by Stimulation Experiments With Econometric Models," *Econometrica*, Apr. 1969, *37*, 333–52.

G. W. Scully, "Business Cycles and Industrial Strike Activity," *J. Bus., Univ. Chicago*, Oct. 1971, *44*, 359–74.

G. J. Stigler and J. K. Kindahl, *The Behavior of Industrial Prices*, New York 1970.

Buyer's Prices, Seller's Prices, and Price Flexibility: Reply

By GEORGE J. STIGLER*

The application by Douglas Bohi and Gerald Scully of spectral analysis to the price data presented in the book written by James Kindahl and myself is unsuccessful whether judged by economic or statistical criteria. The main deficiencies are five:

1. The spectral analysis is applied to the price indexes corrected for trend. This procedure preserves the high serial correlation in the price series even for long lags. The analysis should have been applied to the first differences in the price indexes or to the first differences in their logarithms.[1]

2. Spectral analysis is based upon the assumption that one is dealing with a stationary, Gaussian time-series. The reported high serial correlations are a denial of the stationarity of the price series. The numerous Board of Labor Statistics (*BLS*) price series with few but large movements do not fit the Gaussian assumption and their behavior cannot be characterized by spectral densities.

3. If two time-series have the same spectral density functions, this does *not* assert that the two series resemble each other in the least: two random price series, for example, would have the same spectral densities, but be uncorrelated. Bohi and Scully make no detailed analysis of the coherence (correlation) of the two sets of price series.

4. The *power* of the tests employed to determine whether the *BLS* and National Bureau price series have the same spectral densities is extremely low. The method used by Bohi and Scully to combine the 31 spectra further reduces the power of the tests. It is therefore remarkable that 18 of 64 pairs of price series are determined to be different.

5. The confidence intervals used to judge the significance of the differences in the spectral densities are F-tests, for whose validity it is necessary to assume that the two price series are *independent*. The test is then used to conclude that usually they are not independent.

REFERENCES

D. R. Bohi and G. W. Scully, "Buyer's Prices, Seller's Prices, and Price Flexibility," *Amer. Econ. Rev.*, June 1975, *65*, 517–25.

G. E. P. Box and P. Newbold, "Some Comments on a Paper by Coen, Gomme, and Kendall," *J. Royal Statist. Soc.*, Series A, 1971, *134*, 229–40.

G. J. Stigler and J. K. Kindahl, *The Behavior of Industrial Prices*, New York 1970.

---

* University of Chicago. I am indebted to Stephen M. Stigler for helpful discussions.

[1] For dramatic examples of the effects of routine detrending, see G. E. P. Box and P. Newbold.

ERRATA

Please note corrections and additions to the October 1974 *Directory of Members*

I. Due to editorial errors, the following corrections are noted

Correct Name

Bornhofen, John O. (Listed as Barnhofen, John O.)
Hoffmann, Charles. (Listed as Hoffman, Charles)
López-Mangual, Juan. (Listed as Mangual, Juan López)
Mattart, Fernand. (Listed as Fernand, Mattart)
Sparks, Zoe Ann. (Listed as Sparks, Jr., Jared)
Talmers, William N. (Listed as Talmere, William N.)
Van Roy, Edward. (Listed as Van, Al, Edward)
Vander Weide, James H. (Listed as Weide, James Vander)

Correct Previous and Current Position

Feinn, Barbara A. Cur. Pos. Chief Economist, State of South Carolina.

Harbeson, Robert W. Prev. Pos. Assoc. Prof. 1947–54, Prof., 1954–65, McKinley Prof. of Econs., U. of Ill., 1965–71.

Kaser, Michael. Cur. Pos. Professorial Fellow; Reader in Economics, St. Antony's College, Oxford U.

Correct Degree

Gruen, Edward D. Educ. Ph.D.-elect, Duke U., 1949 (incorrectly given as completed).

Lipsky, John P. Educ. Ph.D. (expected), Stanford U., 1974 (incorrectly given as completed).

Correct Address

Spiro, Michael H. Address Senior Economist, U.S. Dept. of Commerce, Office of the Secretary, Washington, DC 20230. (Mr. Spiro was incorrectly listed as Chair, Dept. of Economics, U. of Pittsburgh.)

Classification by Academic Affiliation

US NAVAL ACADEMY

Annapolis, MD

Fraas, A. G.
Fredland, J. E.
Fredland, J. R.
Gibb, A.
Goodman, R. J.
Little, R. D.
Morris, Jr., C. E.
Whitaker, A. R.

II. Due to computer errors, the following names and biographical sketches were omitted

Name and Address Only

James, Robert G. Address 80 Ludlow Dr., Chappaqua, NY 10514

Killeen, John F. Address Loyola Marymount University, 7101 W. 80th, Los Angeles, CA 90045

Weintraub, Sidney. Address Dept. of Economics, The Wharton School, 3718 Locust Walk, U. of Pa., Philadelphia, PA 19174

Complete Biographies

De Bard, Jr., August A., government; b. Stapleton, Staten Island, N.Y., 1904. **Educ.** B.C.S., New York U., 1929; M.C.S., New York U., 1932; M.A., New York U., 1936. **Fields** 520, 320, 430. **Prev. Pos.** City Mgr., Peoria, Ill., 1960–63; City Mgr., Halifax, Nova Scotia, 1952–60; Fin. Dir., Milford, Conn., 1947–52. **Cur. Pos.** Retired Pres., Fransvea Co. **Address** P.O. Box 283, Laurel, DE 19956

Giersch, Herbert, academic; b. Reichenbach/Eulengeb., 1921. **Educ.** Ph.D., U. Münster, 1948. **Doc. Dis.** Der Ausgleich der Kriegslasten vom Standpunkt sozialer Gerechtigkeit, 1948. **Fields** 400, 130, 300. **Pub.** Allgemeine Wirtschaftspolitik, Grundlagen, Wiesbaden, 1960; Annual Reports 1964–68 of the German "Council of Experts on Economic Development" (co-author), 1965–68; Rationale Wirtschaftspolitik in der pluralistischen Gesellschaft, Rationale Wirtschaftspolitik und Planung in der Wirtschaft von heute, Schriften des Vereins f. Socialpol. N.F. Bd. 45, Berlin, 1967. **Prev. Pos.** Prof. econs., U. of the Saar, Saarbrücken, 1955–69; fellow at the Dev. Centre of the OECD, Paris, 1963–75; member of the German Council of Experts on Econ. Dev., 1964–70. **Cur. Pos.** Dir., Institut für Weltwirtschaft, since 1969, prof. econs., U. Kiel. **Address** Universität Kiel, Dusternbrooker Weg 120, 23 Kiel, Germany

Olsen, Edgar O., academic; b. New Orleans, La., 1942. **Educ.** B.A., Tulane U., 1963; Ph.D., Rice U., 1968. **Doc. Dis.** A Welfare Economic Evaluation of Public Housing, 1968. **Fields** 320, 930, 210. **Pub.** "A Normative Theory of Transfers," Pub. Choice, 1969; "A Competitive Theory of the Housing Mkt.," Amer. Econ. Rev., 1969; "An Econometric Analysis of Rent Control," Jour. of Political Economy, 1972. **Res.** Econ. Effects of Govt. Housing Programs. **Prev. Pos.** Asst. Prof., U. of Va., 1970–72; Econs., Rand Corp., 1968–70; Postdoctoral Fellow, Indiana U., 1967–68. **Cur. Pos.** Assoc. Prof., U. of Va. since

1972. **Address** James Wilson Dept. of Econs., Rouss Hall, Univ. of Virginia, Charlottesville, VA 22901

Recktenwald, Horst Claus, academic; b. Spiesen, 1920. **Educ.** Diplomvolkswirt, U. Mainz, 1949; Dr. rer. pol., 1950; **Doc. Dis.** Steuerinzidenzlehre, 1958 **Fields** 030, 110, 310, 320. **Pub.** Lebensbilder grosser Nationalökonomen, Einführung in die Geschichte der Politischen Ökonomie, 1965; Die Finanzwissenschaft unserer Zeit, 1965; Tax Incidence and Income Redistribution. **Res.** Efficiency and rational decision in Government—Market and Public sector. **Prev. pos.** Chmn., Deutscher Wirtschafts- und Sozialwissenschaftlicher Fakultätentag, 1966–69; Prorektor, prof. of Wirtschafts- und Sozialwissenschaftliche Fakultät der U. Erlangen-Nürnberg, 1968–69; Prof. Wirtschafts- u. Sozialwiss. **Cur. pos.** Fakultät Erlangen-Nürnberg since 1961; Dir., Institut für Wirtschafts- und Finanzpolitik, Friedrich-Alexander-Universität Erlangen-Nürnberg. **Address** Friedrich-Alexander-Universität Erlangen-Nürnberg, Institut für Wirtschafts- und Finanzpolitik, 85 Nürnberg, Germany

Smith, Bernard Joseph, business; b. Liverpool, Eng., 1900. **Educ.** B. Eng., U. Liverpool, 1923; M. Eng., 1926. **Fields** 130, 850, 940. **Pub.** "The Port of San Francisco," Bus. Digest, 1968; "Urban Transportation," Bus. Digest, 1968; "Pollution Hazards in San Francisco Bay," Bus. Digest, 1967. **Prev. Pos.** Corps Engrs., U.S. Army, 1957–65; staff engr. gen. conslt., San Francisco Bay Conservation & Dev. Commn., 1966; cons. engr., U.S. Army. **Cur. Pos.,** cons. engr. **Address** P.O. Box 663, Aptos, CA 95003

We appreciate the time members took to write us about the numerous errors in the Geographical List of Members. Space constraints are such that we cannot print the corrections. It is evident that this section will require special handling in the next edition of the *Directory,* which should be published in the next five years.

Names of Administrative Officers of Departments of Economics have been corrected on the mailing list maintained by the Association.

The Association does not keep corrections and additions on file for future editions of the *Directory.* Please do not send questionnaires to the Association office or to the data service company. Members in good standing at the time the next *Directory* is published will be asked to correct their 1974 listing or submit a new questionnaire.

NOTES

Announcement of Changes in Editorial Policy

The May 1975 issue of the *Review* (*Papers and Proceedings*) contains a financial report by the Secretary-Treasurer for the year 1974, and a projection of operations for 1975. Due to sharp increases in printing and mailing costs, in salaries and office expenses, and to the increased scope of the Association's activities, 1974 expenses outpaced revenues by more than $125,000. This situation will worsen in 1975. While a small fraction of these expenses are accruals to reserves (such as the cost of producing the membership *Directory*), the deficit represents in large part a cash loss which must be met in the short run by borrowing or liquidation of securities and in the long run by economies of operation and increased revenues. Accordingly the following methods of reducing cost or raising revenue have been recommended to the Executive Committee, and accepted at its meeting on March 14, 1975.

a) The discontinuance of free reprints/offprints to authors of material appearing in the *American Economic Review* and *Journal of Economic Literature*. Reprints /offprints will continue to be sold at prices which cover the relevant costs. This change will begin with the December 1975 issues.

b) A differential submission fee for authors who are not members of the American Economic Association. Beginning July 1, 1975, the submission fee for manuscripts submitted to the *American Economic Review* will be fifteen dollars ($15) for members and thirty dollars ($30) for nonmembers.

c) A postage surcharge for mailing of material outside the United States. This surcharge will equal the difference between mailing costs inside and outside the United States. It will be added to new subscription and membership bills January 1, 1976.

d) A page charge for articles, shorter papers, notes, comments, and replies published in the *American Economic Review* and in the *Journal of Economic Literature*. The page charge is payable by the institution or granting agency supporting the research. Payment of this charge is not a prerequisite for publication, nor are authors expected to pay these charges themselves. The charge will be twenty-five dollars ($25) per printed page, and will go into effect with manuscripts accepted after July 1, 1975.

Nominations for AEA Officers

The Electoral College on March 14 chose Lawrence R. Klein as nominee for President-Elect of the American Economic Association in the balloting to be held in the autumn of 1975. Other nominees (chosen by the nominating committee) are: for Vice-President (two to be elected), Barbara R. Bergmann, Harry G. Johnson, Douglass C. North, and Albert E. Rees; for member of the Executive Committee (two to be elected), Robert W. Fogel, Bert G. Hickman, Edmund S. Phelps, and Alice M. Rivlin.

Under a change in the bylaws as described in the *Papers and Proceedings* of this *Review*, May 1971, page 472, additional candidates may be nominated by petition, delivered to the Secretary by August 1, including signatures and addresses of not less than 6 percent of the membership of the Association for the office of President-Elect and not less than 4 percent for each of the other offices. For the purpose of circulating petitions, address labels will be made available by the Secretary at cost.

1976 Nominating Committee of the AEA

In accordance with Section IV, paragraph 2, of the bylaws of the American Economic Association as amended in 1972, President-Elect Franco Modigliani has appointed a Nominating Committee for 1976 consisting of Kenneth Arrow, Chairman; William James Adams; Duran Bell; Marianne Ferber; Jack Hirschleifer; Leonard Rapping; and Vincent Tarascio. Attention of members is called to the part of the bylaw reading, "In addition to appointees chosen by the President-elect, the Committee shall include any other member of the Association nominated by petition including signatures and addresses of not less than 2 percent of the members of the Association, delivered to the Secretary before December 1. No member of the Association may validly petition for more than one nominee for the Committee. The names of the Committee shall be announced to the membership immediately following its appointment and the membership invited to suggest nominees for the various offices to the Committee."

Nicolaus and Associates have been appointed the official travel coordinator for the Dallas meetings of the Allied Social Science Associations. Group air fares may be arranged from your city and for groups as small as fifteen. These fares usually run about 25 percent below regular coach fare. Some examples using current fares subject to change are:

	Coach	Group
New York-Dallas-New York	$236.73	$159.73
Des Moines-Dallas-Des Moines	124.73	99.73
San Francisco-Dallas-San Francisco	252.73	202.73

Flights will be arranged for other cities if enough individuals show interest. Air reservation forms and general information on flights and costs will be mailed to you in October with the preregistration material. In the meantime if you have any questions on group flights, please call Nicolaus and Associates collect at 817-461-1622, or write them at 1201 N. Watson Road, Suite 179, Arlington, Texas 76011.

The German Marshall Fund of the United States, an American private foundation, announces that *German Marshall Fund Common Problems Fellowships* will be available in 1976 to outstanding U.S. scholars whose work is designed to contribute "to the better understanding and resolution of significant contemporary or emerging common problems of industrial societies," particularly in their comparative, international, social, political and economic aspects. Write for information and application form (deadline Oct. 31, 1975) to the German Marshall Fund, 11 Dupont Circle, N.W., Washington, D.C. 20036.

Economists who are *strongly* oriented toward the humanities, who use humanistic methods in their research, and who will be participating in meetings abroad that are concerned with the humanistic aspects of their discipline are eligible to apply for travel grants of the American Council of Learned Societies. Specifically, economists may be eligible if (a) they deal with the history of economic thought or economic history and (b) if their approach is qualitative and descriptive rather than quantitative and statistical. Conferences dealing with the establishment of social policy or legislation are generally ineligible. The deadlines for applications to be received in the office of the American Economic Association are: for meetings scheduled between June and September, February 15; for meetings scheduled between October and January, June 15; for meetings scheduled between February and May, October 15. Application forms may be obtained from Rendigs Fels, Secretary, American Economic Association, 1313 21st Avenue South, Nashville, Tennessee 37212.

The eighty-eighth annual meeting of the American Economic Association will be held in Dallas, Texas, December 28–30, 1975. The Employment Registry and Center will open December 27. The meeting was originally scheduled for October, but has been changed in the interest of attracting greater attendance. The preliminary program of the meeting will appear in the September 1975 issue of this *Review*.

Besides its normal run, *Social Science Quarterly* periodically organizes and publishes special issues dealing with specific topics that are timely and will be of interest to its multidisciplinary audience. Planned for 1976 is a special issue focussing on "Society and Scarcity." It will be devoted to theoretical and empirical inquiries into various aspects of this topic: including human adaptations (or lack of adaptations) to scarcity, equity, and the distribution of income; power and the distribution of income; the distribution of income and social conflict; the international distribution of income and international relations; scarcity and social values (particularly the impact of scarcity on environmental protection and conservation); food and the future; and the energy crisis. Short to medium length manuscripts (12–29 pages) dealing with any aspect of this topic are invited, and should be submitted for consideration before November 1, 1975, to the Editor, *Social Science Quarterly*, University of Texas, Austin, Texas, 78712.

A conference on The Ukraine Within the USSR After World War II: An Economic Balance Sheet will be held at the Harvard Ukrainian Research Institute, Harvard University, 1581–1583 Massachusetts Avenue, Cambridge, Massachusetts, on Friday and Saturday, September 26 and 27, 1975. The conference will consist of three sessions at which eleven papers will be presented and discussed. For further information contact Professor I. S. Koropeckyj, Department of Economics, Temple University, Philadelphia, PA 19122.

The next volume (XIV) of the *Annals of the Ukrainian Academy of Arts and Sciences in the U.S.A.* will be devoted to economic problems. Articles on the current situation of Ukrainian economy, history of Ukrainian economy, the history of economic thought in the Ukraine, and reviews of relevant works are solicited. The deadline for submitting of articles and reviews is the end of 1975. Publication of the issue is planned for the middle of 1976. Professor I. S. Koropeckyj, Department of Economics, Temple University, Philadelphia, PA 19122.

The International Savings Banks Institute, Geneva, Switzerland, has organized an international prize competition on the subject of:

> Indexation of monetary assets and (credit) obligations—pros and cons, taking into account probable effects on banking and the economy; practical possibilities and consequences.

First prize is 15,000 Swiss francs. Second prize is 5,000 Swiss francs. Inquiries and correspondence should be addressed to: International Savings Banks Institute, 1-3 Rue Albert-Gos, 1206 Geneva, Switzerland.

The Urban and Regional Information Systems Association (URISA) is holding its 13th annual conference at the Washington Plaza Hotel in Seattle, Washington, August 24–28, 1975. The theme will be The Role of Information Systems Technology in Community Management.

Urban Economics, a special summer program, will be held at Massachusetts Institute of Technology, August 11–22, 1975, featuring classroom work and distinguished lecturers discussing the most recent theoretical and empirical economic findings in the area of urban economics ranging from housing and transportation to the provisions organization and pricing of government services, etc. For further information, please write to: Director of the Summer Session, Room E19-356, M.I.T.; Cambridge, Massachusetts 02139.

The second annual meeting of the European Finance Association will be held from September 24–26 at the London Business School. For information, write to: Professor Richard Brealey, London Business School, Sussex Place, Regent's Park, London, NW1.

Call for Papers

The purpose of this special issue of *Management Science* on the Leisure Industry is to present a variety of applications of OR/MS techniques to the activities of the leisure-time industry, and to attract additional practitioners into this area of applications. The issue will contain both invited and contributed papers. Papers should deal with any aspect of the leisure-time industry: tourism, travel, recreation, gambling, arts, and lodging. Optimization studies (e.g. optimal utilization of an amusement park, optimal facility mix for private or public facilities such as country clubs, parks, wilderness areas, etc.) are preferred, although statistical analysis will also be considered. Validated applications are encouraged. The length of papers should not exceed 20 journal pages. Short notes will also be considered. All papers will be subject to the normal refereeing process. The issue is being coedited by Shaul P. Ladany, Haifa University, and V. Kerry Smith, State University of New York at Binghamton, Department of Economics, Binghamton, New York 13901. Send abstracts or descriptions of your topic. Completed manuscripts (4+4 copies) will be required by December. All correspondence should be sent to V. K. Smith.

The University of Missouri-Rolla, in cooperation with the Missouri Energy Council is sponsoring the second annual UMR/MEC Energy Conference to be held at the University of Missouri-Rolla, October 7–9, 1975. The program chairpersons for economics are Christopher Garbacz and Vaman Rao, Department of Economics, University of Missouri-Rolla, Missouri 65401.

The Rockefeller and Ford Foundations are continuing for a fifth year a program in support of population policy research in the social sciences. The program is designed to encourage research on the socioeconomic determinants and consequences of population behavior relevant to the process of population policy formulation and implementation. Proposals must be received by July 1, 1975. Awards will be announced by the end of December 1975. All inquiries and proposals should be sent directly to The Rockefeller-Ford Program for Population Policy Research, The Rockefeller Foundation, P.O. Box 4516, Grand Central Station, New York, New York 10017.

Deaths

Procter Thomson, professor of economics, Claremont Men's College, Jan. 26, 1975.

Retirements

Willard D. Arant, economist, Swift and Company, Nov. 1975.

George W. Zinke, department of economics, University of Colorado, Aug. 1973.

Visiting Foreign Scholars

B. Curtis Eaton, Queen's University: visiting scholar, department of economics, University of Colorado, 1974–75.

Richard Lipsey, Queen's University: visiting scholar, department of economics, University of Colorado, 1974–75.

Motoaki Moriya, Kansai University, Asaka, Japan: visiting scholar, department of economics, University of Colorado, Jan. 1975.

Promotions

Roy C. Anderson: associate professor, department of economics, California State University, Long Beach.

G. Knight Boyer: associate professor, department of economics, geography and management, U.S. Air Force Academy, July 1, 1974.

John S. Brush: associate professor, department of economics, geography and management, U.S. Air Force Academy, July 1, 1973.

James F. Crook: associate professor of economics and statistics, Winthrop College, Aug. 1974.

James J. Doyle: assistant professor, department of economics, John Carroll University.

Fred R. Glahe: professor, department of economics, University of Colorado, fall 1973.

Arthur Kraft: professor, department of economics, Ohio University, June 1975.

Mary Lou Larmore: associate professor, department of economics, California State University, Long Beach.

Perry W. Polk: assistant professor, department of economics, geography and management, U.S. Air Force Academy, July 1, 1974.

Robert F. Rooney: professor, department of economics, California State University, Long Beach.

Larry D. Singell: professor, department of economics, University of Colorado, June 1973.

Andrew Stern: associate professor, department of economics, California State University, Long Beach.

Robert L. Taylor: assistant professor, department of economics, geography and management, U.S. Air Force Academy, July 1, 1973.

Murray Yanowitch: professor, department of economics, Hofstra University, Dec. 16, 1974.

Administrative Appointments

Edward L. Claiborn: professor; head, department of economics, geography and management, U.S. Air Force Academy, Aug. 1, 1974.

Simeon J. Crowther: chairman, department of economics, California State University, Long Beach.

Herbert C. Morton, Bureau of Labor Statistics: director of public affairs, Resources for the Future, Feb. 1975.

Charles E. Staley: acting chairman, department of economics, State University of New York, Stony Brook, Feb. 4, 1975.

Appointments

James T. Bennett, George Washington University: associate professor of economics, George Mason University, Sept. 1975.

Claire Bombardier: research fellow, National Bureau of Economic Research, Jan. 6, 1975.

Yale N. Braunstein, Stanford University: research associate in economics, New York University, Sept. 1974.

E. J. Devine: associate director, Defense Manpower Commission, Nov. 1974.

Thomas P. Drinka: research associate, department of economics, Iowa State University, Nov. 1, 1974.

Margo Faier: economist, Banking Studies Division, Federal Reserve Bank of New York, Oct. 10, 1974.

Francis Lea Kaiser: research associate, department of economics, Iowa State University, Oct. 3, 1974.

Richard D. Karfunkle: chief economist; director, economics group, Lehman Brothers, Inc., Jan. 1975.

Avram Kisselgoff: senior research associate, National Bureau of Economic Research, Nov. 1, 1974.

John F. Kyle, Exxon Corporation: assistant professor of economics, New York University, Sept. 1974.

Ludwig Lachmann, University of Johannesburg, South Africa: visiting professor of economics, New York University, Feb. 3, 1975.

Richard A. Levins: research associate, department of economics, Iowa State University, Dec. 20, 1974.

Treena K. Levins: research associate, department of economics, Iowa State University, Dec. 20, 1974.

Ann R. Markusen: assistant professor, department of economics, University of Colorado, Spring 1974.

A. James Meigs, Argus Research Corporation: professor of monetary theory and banking practices, Claremont Men's College, Feb. 1975.

Warren E. Moskowitz, Exxon Corporation: assistant professor of economics, New York University, Sept. 1974.

Gordian Nworah: economist, Banking Studies Division, Federal Reserve Bank of New York, July 25, 1974.

Craig W. O'Riley: research associate, department of economics, Iowa State University, Nov. 16, 1974.

Joel Popkin: senior research associate, National Bureau of Economic Research, Nov. 1, 1974.

William Rutledge: economist, Banking Studies Division, Federal Reserve Bank of New York, Sept. 5, 1974.

Thomas A. Wolf, International Institute of Management, Berlin, Germany: assistant professor, department of economics, Ohio State University.

Edward N. Wolff, Yale University: assistant professor of economics, New York University, Sept. 1974; research associate, National Bureau of Economic Research, Nov. 1, 1974.

Leaves for Special Appointments

Elbert V. Bowden, State University of New York, College at Fredonia: team leader, United Nations regional economic planning team, Fiji Islands, 1975–76.

David E. R. Gay, University of Arkansas: visiting scholar, Institute on Soviet and East European Studies, University of Glasgow, May 1975.

Leslie E. Grayson, University of Virginia: Ford Foundation award, visiting professor, European Graduate Institute of Business Administration, 1974–75.

Karl D. Gregory, Oakland University: visiting faculty professor, State University of New York, Buffalo, Jan. 6, 1975.

Robert L. Harlow, College of the Holy Cross: international economist, Office of International Monetary Affairs, U.S. Treasury Department, Jan. 6, 1975.

Toussaint Hočevar, University of New Orleans: Fulbright-Hays senior researcher, University of Ljubljana, Slovenia, Feb.–June 1975.

Arthur Kraft, Ohio University: Sears-AACSB Federal Faculty Fellow, Office of Research and Statistics, Social Security Administration.

Dudley G. Luckett, Iowa State University: visiting professor, department of economics, University of Tulsa, Feb.-May 1975.

Alexander Melamid, New York University: visiting professor of economic geography, University of Sydney, Australia.

Resignations

David Calvin Gogerty, The Rand Corporation, Oct. 1974.

NOTE TO DEPARTMENTAL SECRETARIES AND EXECUTIVE OFFICERS

When sending information to the *Review* for inclusion in the Notes Section, please use the following style:
A. Please use the following categories:

1—Deaths
2—Retirements
3—Foreign Scholars (visiting the USA or Canada)
4—Promotions
5—Administrative Appointments
6—New Appointments
7—Leaves for Special Appointments (NOT Sabbaticals)
8—Resignations
9—Miscellaneous

B. Please give the name of the individual (SMITH, John W.), his present place of employment or enrollment: his new title (if any), his next place of employment (if known or if changed), and the date at which the change will occur.

C. Type each item on a separate 3×5 card, and please do not send public relations releases.

D. The closing dates for each issue are as follows: *March*, November 1; *June*, February 1; *September*, May 1; *December*, August 1.

This announcement supersedes and replaces a letter which was sent annually from the managing editor's office. All items and information should be sent to the Assistant Editor, *American Economic Review*, Box Q, Brown University, Providence, Rhode Island 02912.

NOTICE TO ALL GRADUATE DEPARTMENTS

The December 1975 issue of the *Review* will carry the seventy-second list of doctoral dissertations in political economy in American universities and colleges. The list will specify doctoral degrees conferred during the academic year terminating June 1975. This announcement is an invitation to send us information for the preparation of the list. This announcement supersedes and replaces a letter which was sent annually from the managing editor's office.

The *Review* will publish in its December 1975 issue the names of those who will have been awarded the doctoral degree since June 1974 and the titles of their dissertations. Dissertation abstracts will no longer be published, as these are published elsewhere.

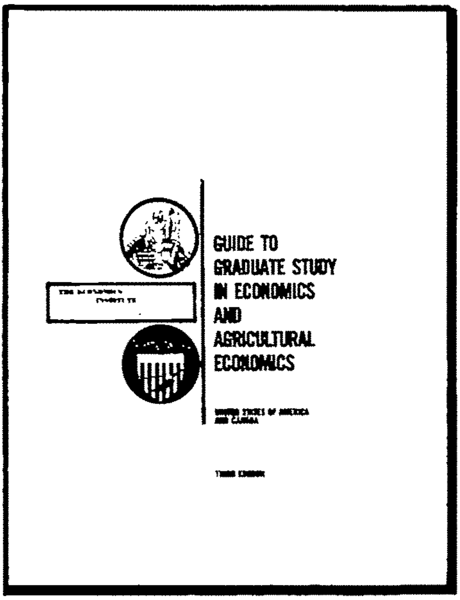
By June 30, please send us this information on 3×5 cards, conforming to the style shown below, one card for each individual. Please indicate by a classification number in the right-hand corner the field in which the thesis should be classified. The classification system is that used by the *Journal of Economic Literature* and printed in every issue.

JEL Classification No. ________
Name: LAST NAME IN CAPS: First Name, Initial
Institution Granting Degree: ________
Degree Conferred (Ph.D. or D.B.A.) ________ Year ________
Dissertation Title: ________

When degrees in economics are awarded under different names, such as Business Administration, Public Administration, or Industrial Relations, candidates in these fields whose training has been *primarily in economics* should be included.

All items and information should be sent to the Assistant Editor, *American Economic Review*, Box Q, Brown University, Providence, Rhode Island 02912.

Announcing...



GUIDE TO GRADUATE STUDY IN ECONOMICS AND AGRICULTURAL ECONOMICS

Guide to Graduate Study in Economics and Agricultural Economics

in the United States of America and Canada

THIRD EDITION

Designed to provide students anticipating graduate study in economics and agricultural economics, and their advisors, with information on available graduate training programs.

Includes descriptions of 236 graduate programs, supplemented by comparative data and information for prospective students, domestic and foreign.

Published by the Economics Institute under the auspices of the American Economic Association, and the American Agricultural Economics Association.

Distributed by:
RICHARD D. IRWIN INC.,
1818 Ridge Road,
Homewood, Illinois 60430

PRICE: $7.50 per copy

DETENTE AND DOLLARS:

Doing Business with the Soviets

Marshall I. Goldman

This indispensable book by a leading expert in East-West trade, provides the American businessman with the perspective and information he needs to decide whether he too should enter the billion-dollar Soviet market and, if so, how he can do it successfully. Dr. Goldman draws important lessons from his detailed inside account of both costly failures (the great American grain disaster of 1972) and success stories (Armand Hammer and the Pullman Corporation) in doing business with the Soviet Union.
Due June $10.00

HOUSING AND THE MONEY MARKETS

Roger Starr

Here is the first systematic introduction to the money markets as they relate to housing. Now Administrator of New York City's Housing and Development Administration, Mr. Starr describes the various institutions that extend mortgages and shows how they finance housing construction. He explores the relationship between lending institutions and the various government agencies such as the FHA, GINNY MAE, and FANNY MAE that have been set up for channeling savings into mortgages. "Timely as well as valuable. . . . An excellent primer and a well reasoned policy text for specialists and government officials."—LOUIS WINNICK.
Just published $10.95

ANARCHY, STATE AND UTOPIA

Robert Nozick

In the most challenging defense of property rights and free enterprise in modern times, Professor Nozick poses, as *Forbes* puts it, "nothing less than a challenge to a central feature of the moral landscape: the state as everybody's mother and father."—*Forbes.* "No contemporary philosopher possesses a more imaginative mind, broader interests, or greater dialectic abilities."—*Harpers.* Published $12.95

WHO SHALL LIVE?

Health, Economics, and Social Choice

Victor R. Fuchs

"A short lucid, uncomplicated explanation of the economics of health care. . . . No book has demonstrated a better grasp of the *social* context of health and health care, of the idea that central social changes, rather than mere technical tinkering with the system is at issue. Most useful."
—*The New York Times Book Review.*
Just published $8.95

CRISIS IN KEYNESIAN ECONOMICS

Sir John Hicks

Just published $5.95

NEW. SIGNIFICANT.
BASIC
BASIC BOOKS INC.
10 EAST 53RD ST, NEW YORK 10022

Join the American Economic Association

Membership includes:

- a subscription to both *The American Economic Review* (quarterly) plus May, *Papers and Proceedings* and the *Journal of Economic Literature* (quarterly)
- annual ballots and nomination forms for elected officers of the Association
- preregistration material and announcement of the annual meeting of the Allied Social Science Associations
- announcements of special interest to members, such as the life insurance program started in 1973; new book offerings, etc.

Application for Membership

Date ____________

Please enroll me as a member of the American Economic Association. 1975 annual rates: Regular—$23; Family, without publications—$2.30; Junior, *registered student—$11.50/payable yearly; for a maximum of three years; Subscribing—$27.50; Contributing—$57.50; Life—$460. All dues payable in U.S. currency. Memberships are not entered until payment has been received. (Under the terms of a resolution passed at the annual meeting of members of the American Economic Association on Dec. 29, 1973. $18.40 of the annual membership dues is considered as payment for the **AER** and **JEL**; the remainder is considered a contribution to exempt functions of the Association.)

Name in full ____________

Address ____________

____

City State/Country Zip/Post Code

Endorsed by (AEA member) ____________

Below for Junior Members Only

* I certify that the person named above is enrolled as a student at ____________

____

NOTE: Payment and certification must be submitted yearly.

____

Authorized Signature

Please send with remittance to: American Economic Association
1313 21st Avenue South
Nashville, Tennessee 37212

A devastating analysis of the mysterious corporate world—operated by "World Managers" whose loyalties are to their companies rather than to any nation, including their own.

GLOBAL REACH

"Clearly the definitive book on multinationals." —Houston Chronicle

"Fascinating." —New York Times Book Review

"A superb book." —John Kenneth Galbraith

"Rewarding, readable and provocative." —Business Week

$11.95

The Power of the Multinational Corporations

RICHARD J. BARNET & RONALD E. MÜLLER

THE PERMANENT WAR ECONOMY

American Capitalism in Decline

by Seymour Melman,

author of *Pentagon Capitalism*

The powerful and controversial book which challenges the widely held belief that defense spending has been good for the American economy.

"A reasoned, documented account of how and why the country is falling apart." —San Francisco Chronicle

$9.95

SIMON AND SCHUSTER

STAFF PAPERS

Published by the International Monetary Fund three times a year with summaries in English, French, and Spanish

Vol. XXII March 1975 No. 1

The Role of Incomes Policy in Industrial Countries Since World War II ....Anne Romanis Braun

The Impact of U.S. Controls on Capital Outflows on the U.S. Balance of Payments: An Exploratory Study ....John Hewson and Eisuke Sakakibara

Fiscal Incidence Studies in Developing Countries: Survey and Critique ....Luc De Wulf

Financial Intermediation, Savings Mobilization, and Entrepreneurial Development: The African Experience ....Rattan J. Bhatia and Deena R. Khatkhate

Describing External Debt Situations: A Roll-Over Approach ....Pierre Dhonte

Tax Ratios and Tax Effort in Developing Countries ....
..................................Raja J. Chelliah, Hessel J. Baas, and Margaret R. Kelly

The Fund Agreement in the Courts—XI ....Joseph Gold

Staff Papers is one of four subscription periodicals published by the Fund: *Staff Papers* ($6 a year or $2.50 for a single copy), *Balance of Payments Yearbook* ($7.50), *Direction of Trade* ($10), and *International Financial Statistics* ($20). Special rates to university libraries, faculty members, and students: $3 a year for each of the first three titles (single copies of *Staff Papers* are $1) and $5 a year for *International Financial Statistics*, or all four periodicals for $12 a year. Subscriptions in other currencies accepted.

Address inquiries to The Secretary, International Monetary Fund, Washington, D.C. 20431, U.S.A.

field studies and fields of study

Extension in the Andes
by E. B. Rice
$17.50

The Andean Group:
A Case Study in Economic Integration among Developing Countries
by David Morawetz
$18.50

Agriculture in the Postbellum South:
The Economics of Production and Supply
by Stephen J. DeCanio
$15.00

Planning for an Urban World:
The Design of Resource-Conserving Cities
by Richard L. Meier
$27.50

Toward an Economic Theory of Income Distribution
by Alan S. Blinder
$12.50

Selected Readings in Macroeconomics and Capital Theory from ***Econometrica***
edited by David Cass and Lionel W. McKenzie
$22.50

Economics and Public Policy:
The Automobile Pollution Case
by Donald N. Dewees
$18.50

Airport Economic Planning
edited by George P. Howard
$18.95

The MIT Press
Massachusetts Institute of Technology
Cambridge, Massachusetts 02142

Black-White Income Differentials

Empirical Studies and Policy Implications

by STANLEY H. MASTERS

A Volume in the INSTITUTE FOR RESEARCH ON POVERTY MONOGRAPH Series

This book investigates why urban blacks have lower incomes than whites in similar occupational positions. Professor Masters presents statistical tests of three widely held hypotheses of urban black discrimination—1) southern migrants cause the low income of blacks in northern cities, 2) housing segregation causes significant problems for urban black males, 3) differences in productivity are more important than labor-market discrimination. The book presents data showing that the first two hypotheses are false and the validity of the third depends upon assumptions about the reliability of years of schooling or test-score performance as measures of productivity. The author concludes that much of the racial differential in earnings results from labor-market discrimination.

CONTENTS: Economic Theories of Discrimination. A Review of Empirical Studies of Discrimination Against Blacks. The Effect of Migration on Black Incomes. The Effect of Housing Segregation on Black-White Income Differentials. Estimating the Relative Importance of Labor-Market Discrimination Versus Differences in Productivity Between Blacks and Whites. Alternative Policy Perspectives. Policy Recommendations.

1975, 204 pp., $12.50/£6.00

Stochastic Optimization Models in Finance

edited by W. T. ZIEMBA and R. G. VICKSON

A Volume in the ECONOMIC THEORY AND MATHEMATICAL ECONOMICS Series

This book fills the need for a thorough account of the mathematical and economic theory relevant to stochastic optimization of models of financial problems that involve uncertainty. Almost every article presents a significant methodological advance of lasting interest for teaching and research in finance.

The book is divided into five parts: Mathematical Tools; Qualitative Economic Results; Static Portfolio Selection Models; Dynamic Models Reducible to Static Models; and Dynamic Models; each part opens with a detailed, cohesive introduction and includes both computational and review, and mind-expanding exercises, to test and widen understanding.

1975, 744 pp., $24.50/£11.75

Prices subject to change without notice.

Academic Press

A Subsidiary of Harcourt Brace Jovanovich, Publishers
111 Fifth Avenue, New York, New York 10003
24-28 Oval Road, London NW1 7DX

RICHARD D. MALLON and
JUAN V. SOURROUILLE

Economic Policymaking in a Conflict Society

The Argentine Case

Argentina is a fascinating and baffling case to scholars of economic development. It has rich agricultural resources, a fully monetized economy, a domestic manufacturing sector that occupies a large share of the active labor force, a relatively high level of literacy, and other attributes that resemble a European nation more than a developing country. At the same time, Argentina has found it difficult to develop the institutions of a modern nation-state and to sustain a satisfactory rate of economic growth. This book is a new and vigorous attempt to explain the Argentinian paradoxes. $16.00t

Harvard University Press

79 Garden Street, Cambridge, Massachusetts 02138

ESSAYS ON JOHN MAYNARD KEYNES

Edited and compiled by MILO KEYNES

A new collection of essays revealing little-known facets of Keynes' life and interests and presenting new interpretations of his economic theories. Among the distinguished contributors are: Austin Robinson; Harry G. Johnson; Joan Robinson; John Kenneth Galbraith; and R. B. Braithwaite. The book includes a chronology, a bibliography and an index, and is illustrated with photographs. $16.50

INTERNATIONAL TRADE AND FINANCE

Frontiers for Research

PETER B. KENEN

This volume, based on a conference at the International Finance Section of Princeton, addresses itself to three questions: What have we learned from recent empirical research? What are the main gaps in present knowledge? How should we go about filling those gaps? Here you will find original articles by most of the major figures in the field including R. M. Stern, W. M. Corden, C. Diaz-Alejandro, B. Balassa, S. P. Magee, J. Helliwell, B. J. Cohen, L. R. Klein, and R. E. Caves. There is also a concluding *Overview* by Kindleberger, Chenery, Bhagwati, R. N. Cooper, W. Schmidt, M. Gilbert, and H. G. Johnson. $19.95

Other Cambridge Books for Economists

INVESTMENT DECISIONS IN THE NATIONALISED FUEL INDUSTRIES

R. W. BATES and N. M. FRASER $14.95

UNEMPLOYMENT, VACANCIES AND THE RATE OF CHANGE OF EARNINGS

A. E. WEBB and R. WEEDEN $7.95

COMPANY FINANCE AND THE CAPITAL MARKET

A Study of the Effects of Firm Size

E. W. DAVIS and K. A. YEOMANS

DAEO Paper 39 Cloth $13.95 Paper $5.95

UNFAIR DISMISSAL

D. JACKSON

DAE Papers in Industrial Relations and Labour 1 Cloth $8.95 Paper $4.50

The Economic Journal

the quarterly journal of the
Royal Economic Society

is now published by Cambridge University Press

The Economic Journal is devoted to the study of all areas of economics, both theoretical and applied, and represents all schools of thought. The journal publishes original articles, surveys and notes of recent research, and an extensive section of book reviews and signed notes on new books.

The Economic Journal is edited by W. B. Reddaway, Professor of Political Economy; D. G. Champernowne, Professor of Economics; and Phyllis Deane, Reader in Economic History, all of Cambridge University.

Contents of Volume 85, Part 337, March 1975

Articles

Ralph Turvey: A simple analysis of optimal fares on scheduled transport service

R. E. Rowthorn: What remains of Kaldor's Law?

Jay H. Levin: Monetary policy and the crawling peg

Mervyn A. King: The U.K. profits crisis: myth or reality?

D. W. Verry and P. R. G. Layard: Cost functions for university teaching and research

Andrew Cosh: The remuneration of chief executives in the United Kingdom

Margo Thorning: Cyclical fluctuations in prices and output in the United States, 1920-1970

James A. Trevithick: Keynes, inflation and money illusion

Ian Steedman: Positive profits with negative surplus value

Notes and Memoranda

Reviews and Notes on New Books

The Economic Journal is published in one volume of four parts annually in March, June, September and December.

Volume price by subscription $36.00.
Single parts $12.00.

Cambridge
University Press
32 East 57th Street,
New York, N.Y. 10022

Begin making plans to attend the

Eighty-eighth
Annual Meeting of The American Economic Association

to be held in

DALLAS, TEXAS

Sunday, Monday, Tuesday

DECEMBER 28-30, 1975

The Employment Center opens Saturday, December 27.
This is a change from the October dates previously announced.

See the Notes section of the September, AER for the preliminary program.

The 1976 meeting will be held in Atlantic City, September 16-18.

Bank Management and Portfolio Behavior

Donald D. Hester and James L. Pierce

Based on a rich body of confidential data on 400 New England banks giving deposits and asset holdings quarter by quarter for a four-year period, this book provides a microeconometric analysis of portfolio behavior and earnings by commercial and mutual savings banks. Cowles Foundation Monograph 25 $20.00

Agricultural Research and Productivity

Robert E. Evenson and Yoav Kislev

The first effort to probe deeply and systematically into the role of agricultural research in producing the type of technology change that can substantially affect food production and the global food problem.

"Makes a major contribution to our understanding of the role of agricultural research in economic development."—D. Gale Johnson

A Publication of the Economic Growth Center, Yale University, in association with the World Bank Cloth $12.50 Paper $3.95

Multinational Firms and Asian Exports

Benjamin I. Cohen

The conventional wisdom holds that developing countries benefit when foreign firms establish plants that produce and export manufactured goods. This monograph by Benjamin Cohen is a searching analysis of that assumption, which the author concludes is incorrect. A Publication of the Economic Growth Center, Yale University $12.50

Colonial Development

An Econometric Study

Thomas B. Birnberg and Stephen A. Resnick

This book uses an econometric model to examine the process of economic development for ten Asian, African, and Latin American countries. It shows how they were transformed in a similar pattern from about the start of the twentieth century to the outbreak of World War II. A Publication of the Economic Growth Center, Yale University $20.00

Now available in paper

Opening Up the Suburbs

An Urban Strategy for America Anthony Downs Cloth $9.50 Paper $2.95

Yale University Press New Haven and London

ADAM SMITH:

Man of Letters and Economist

by CLYDE E. DANKERT

author of Contemporary Unionism in the United States

This book paints a fascinating portrait of one of the most significant figures in the intellectual history of Western man, with new insights on present-day relevance. The literary, as well as the economic, aspects of Smith's career are emphasized in Prof. Dankert's collection of essays. This is stimulating reading for both the specialist and non-specialist. **$10.00**

Published by
EXPOSITION PRESS, INC.
900 South Oyster Bay Road, Hicksville, New York 11801

(When ordering by mail add 50c to cover postage and handling. NYS residents add applicable sales tax)

"... I find the Schiller manuscript excellent. It is extremely well written. It is interesting. It is relevant. It is therefore likely to excite the average undergraduate student."

Walter Adams
Michigan State University

about —

THE ECONOMY

by Bradley R. Schiller

For further information about **The Economy,** by Bradley R. Schiller, write to **Robert Jordan, Dept. J13, College Division, Prentice-Hall, Englewood Cliffs, N.J. 07632**

PRENTICE-HALL

Nicolaus & Associates, Inc.

has been appointed
official travel coordinator
for the meetings of the
Allied Social Science Associations
in Dallas, Texas
December 28-30, 1975

The Employment Center opens Saturday, December 27.

Group air fares may be available from your city and for groups as small as fifteen. These fares usually run about 25% below regular coach fares. For example current fares from three cities:

	Coach	*Group*
New York-Dallas-New York	236.73	159.73
Des Moines-Dallas-Des Moines	124.73	99.73
San Francisco-Dallas-San Francisco	252.73	202.73

Air reservation forms and general information on flights and costs will be mailed to you in October with the pre-registration material. In the meantime if you have any questions on group flights, please call Nicolaus & Associates collect at area code 817-461-1622 or write them at 1201 N. Watson Road, Suite 179,
Arlington, Texas 76011.

The Journal of
LAW & ECONOMICS

The contents of Volume XVII, Number 2, October, 1974 were:

Kenneth W. Dam: The Evolution of North Sea Licensing Policy in Britain and Norway

Reuben A. Kessel: Transfused Blood, Serum Hepatitis, and The Coase Theorem

Paul L. Joskow: Inflation and Environmental Concern: Structural Change in the Process of Public Utility Price Regulation

William Breit and Kenneth G. Elzinga: Antitrust Enforcement and Economic Efficiency: The Uneasy Case for Treble Damages

R. H. Coase: The Lighthouse in Economics

Basil Yamey: Monopolistic Price Discrimination and Economic Welfare

Daniel K. Benjamin and Roger C. Kormendi: The Interrelationship Between Markets for New and Used Durable Goods

Arlene Holen and Stanley A. Horwitz: The Effect of Unemployment Insurance and Eligibility Enforcement on Unemployment

Patrick C. Mann: User Power and Electricity Rates

Linda N. Edwards and Franklin R. Edwards: Measuring the Effectiveness of Regulation: The Case of Bank Entry Regulation

Victor P. Goldberg: Institutional Change and the Quasi-Invisible Hand

R. H. Coase: The Choice of the Institutional Framework: A Comment

The Journal of Law and Economics is published by the University of Chicago Law School. It is published twice a year, in April and October. The annual subscription rate is $9. The price of individual issues is $5. The subscription rate for three years is $23 and for five years $30. Volumes 1 through X were issued annually in October; all these issues are available for purchase at the price of $5 per volume. Volumes XI through XVI are available for purchase at the price of $10 per volume.

Send order, together with check made payable to *The Journal of Law and Economics*, to:

Publications Secretary
University of Chicago Law School
Journal of Law and Economics
1111 East 60th Street, Chicago, Illinois 60637

THE ENERGY CRISIS AND U.S. FOREIGN POLICY
Edited by Joseph S. Szyliowicz and Bard O'Neill

Focuses on the general energy situation in the United States, the economic dimensions of the petroleum trade, reasons for the current crisis, and regional political relationships in the Middle East. Discusses U.S. foreign relations and outlines policy options for the United States.
Apr., 1975 280 pp. ISBN 0-275-09040-X $18.50
PSS Student Edition: ISBN 0-275-88840-1 $6.95

MIDDLE EAST OIL AND U.S. FOREIGN POLICY
With Special Reference to the U.S. Energy Crisis
Shoshana Klebanoff

A survey of U.S. involvement in Middle East oil from 1919 to the present. Assesses the impact of external influences on U.S. policy, finding the locus of decision-making power in U.S. oil policy is centered largely in the private sector.
1974 304 pp. TC 0853 ISBN 0-275-28862-5 $16.50

THE SAUDI ARABIAN ECONOMY
Ramon Knauerhase

A sectoral analysis of Saudi Arabia's economy and development in the context of geography, resources, religion, social systems, and external relations (oil companies). Evaluates the government sector — the social infrastructure, budget, and monetary, revenue, and future planning systems.
Sept., 1975 ca. 350 pp. ISBN 0-275-09000-0 ca. $20.00

PRAEGER SPECIAL STUDIES

THE UNITED STATES AND INTERNATIONAL OIL
A Report for the Federal Energy Administration on U.S. Firms and Government Policy
Robert B. Krueger

Examines the feasibility of an increased U.S. government role in international petroleum affairs to counter the threat posed by the international energy situation to the national interest. Analyzes U.S. policy options in terms of legal, political, and economic factors. Funded by the FEA in response to issues raised by the 1973-74 Arab oil embargo.
Apr., 1975 775 pp. (incl. appendixes)
ISBN 0-275-05300-8 $45.00
PSS Student Edition: 384 pp. (no appendixes)
ISBN 0-275-89340-5 $8.50

Please use TC or ISBN numbers when ordering from:
praeger
111 Fourth Ave.
New York, N.Y. 10003

China's Modern Economy in Historical Perspective

Dwight H. Perkins, Editor. This collection of papers examines the extent to which economic developments after 1949 depended on features unique to China and to Chinese history, as well as on the socialist reorganization of Chinese society. The contributors are *Kang Chao, Robert F. Dernberger, Mark Elvin, John S. H. Fei, Ramon Myers, Dwight H. Perkins, Thomas G. Rawski, Carl Riskin,* and *Peter Schran.* $13.85

French Economic Growth

J.-J. Carré, P. Dubois, & E. Malinvaud. This major attempt to understand and analyze the phenomenal postwar growth of the French economy evaluates the labor force, productivity, investment, capital, and the industrial structure, and examines such non-physical factors as aggregate demand, inflation, economic planning, and foreign trade. *Studies of Economic Growth in Industrialized Countries.* July. About $20.00

Stanford University Press

JOHNS HOPKINS

New this spring

The Japanese Economy in International Perspective

edited by Isaiah Frank

In an effort to improve Western understanding of Japanese economic institutions and policies, these papers illuminate the major aspects of the Japanese economy which affect its external relations. Both American and Japanese contributors cover a variety of topics, ranging from the economic implications of changing Japanese and American security roles in Asia to Japan as host to the international corporation. *Published with the Committee for Economic Development*

$16.00 hardcover, $3.95 paperback

Yugoslavia

Development with Decentralization
World Bank Country Economic Report

In an analysis of the development experience and problems of the Yugoslav economy, this report shows that the success with which Yugoslavia has combined rapid economic growth and fundamental changes over a long period of time speaks for the basic soundness of its mode of economic development. *Published for the International Bank for Reconstruction and Development*

$18.50 hardcover, $8.50 paperback

* * * * * *

The Mystical World of Indonesia

Culture and Economic Development in Conflict

Allen M. Sievers

"Sievers . . . has written, in this full-bodied study of the culture and economics of the varied, attractive, potentially rich group of islands that make up modern Indonesia, a valuable . . . study. . . . The Indonesia people he sees as rooted in nonmaterial satisfaction hence his use of 'mystical' to characterize their unique culture which is likely to resist 'development' in the Western sense for a long time." — *Publishers Weekly*

January 1975, $14.00 hardcover

The JOHNS HOPKINS University Press
Baltimore, Maryland 21218

Educated planning today leads to higher profits tomorrow.

1. ANALYSIS AND CONTROL OF DYNAMIC ECONOMIC SYSTEMS
By **Gregory C. Chow,** Princeton University

Presents a set of related techniques for analyzing the properties of dynamic stochastic models in economics and for applying these models in the determination of quantitative economic policy.

1975 316 pages $19.95

2. LOCATIONAL DYNAMICS OF MANUFACTURING ACTIVITY
Edited by **Lyndhurst Collins,** University of Edinburgh and **David F. Walker,** University of Waterloo, Canada

Original essays interpreting changing patterns of industrial location in terms of historical and behavioral perspectives as well as contemporary theories and models.

1975 402 pages $27.50

3. SOCIAL INFORMATION PROCESSING AND STATISTICAL SYSTEMS—CHANGE AND REFORM
By **Edgar S. Dunn, Jr.,** Resources for the Future, Inc., Washington, D.C.

This book analyzes the ways social data for management, public policy, and social science are acquired and processed and discusses the dilemmas this poses for those seeking to reconcile the need for individual freedom with the equally compelling need for social order.

1974 246 pages $12.50

4. SOCIAL INDICATORS AND SOCIAL THEORY
Elements of an Operational System
By **Karl A. Fox,** Iowa State University

Integrates theory, methods, and data across several social sciences, considering time budgets, behavioral setting, personality structure, occupational structures and many other areas to provide a rich theoretical perspective.

1974 328 pages $14.95

5. SPATIAL PERSPECTIVES ON INDUSTRIAL ORGANIZATION AND DECISION-MAKING
Edited by **F. E. Ian Hamilton,** University of London

Original research studies by well-known authorities focusing upon locational behavior by industrial organizations. Stresses types of industrial organization and decision-making behavior and their impact on industrial location patterns.

1974 533 pages $29.95

6. DOLLARS FOR NATIONAL GOALS
Looking Ahead to 1980
By **Leonard A. Lecht,** National Planning Association

Analyzes the changes in spending patterns during the 1960s, showing how these changes will influence a series of national goals and problems over the next decade.

1974 295 pages $17.95

7. STATISTICAL SAMPLING FOR AUDIT AND CONTROL
By **T. W. McRae,** The University of Bradford Management Centre

This book enables accountants to apply statistical sampling methods in auditing and financial control. It will help reduce the high set up and learning costs that statistical sampling incurs.

1974 279 pages $19.95

Available at your bookstore or mail this coupon to:
WILEY-INTERSCIENCE,
Dept. 527, Box 4569,
Grand Central Station,
New York, N.Y. 10017
In Canada: 22 Worcester Road, Rexdale, Ontario
A 4898-WI

Mail coupon to—
WILEY-INTERSCIENCE, Dept. 527, Box 4569,
Grand Central Station, New York, N.Y. 10017

Please send me the book(s) I have checked below to read and use free for 10 days. At the end of that time, if I am satisfied with my order and want to keep it, I will send you the amount indicated for each book received plus postage and handling. Otherwise, I will return the order and owe nothing.

☐ 1 (1 15616-7)
☐ 2 (1 16582-4)
☐ 3 (1 22747-1)
☐ 4 (1 27060-1)
☐ 5 (1 34715-9)
☐ 6 (1 52095-0)
☐ 7 (1 58991-8)
☐ My check (money order) for $______ is enclosed.
☐ Please bill me. (Restricted to continental U.S.)
☐ Please send me a list of local bookstores carrying your titles.

Name________________________

Affiliation________________________

City____________ State________ Zip________

Please add state and local taxes where applicable.
Prices subject to change without notice. 092 A 4898-WI-1

Job Openings for Economists

A bimonthly listing of job vacancies

Available only to AEA members and to organizations and departments of economics who agree to list their openings.

I would like to subscribe to *Job Openings for Economists* and enclose a check for one year.

1975 Subscription Rates—U.S.A., Canada, and Mexico (first class): $6.00 junior member, $12.00 others; foreign airmail: $12.00 junior member, $18.00 others.

Name ____

Address ____

____

Check one:

☐ I am a member of the American Economic Association.

☐ I am not a member but would like to become one. My application is enclosed.

☐ (For organizations) We agree to list our vacancies in JOE.

THE AMERICAN ECONOMIC ASSOCIATION
1313 21st Avenue South
Nashville, Tennessee 37212

New from Columbia

EMPLOYMENT IN DEVELOPING NATIONS

EDGAR O. EDWARDS, Editor

This book presents the findings of three seminars convened to study the problems of employment in developing countries. The seminars, dealing with economic policy, technology, and rural development, were sponsored by the Ford Foundation and were designed to promote understanding for the alleviation of these problems.

cloth, $15.00; paper, $7.50

SEX, DISCRIMINATION, AND THE DIVISION OF LABOR

CYNTHIA B. LLOYD, Editor

This collection of sixteen essays looks at the role of women in the labor market, touching on the areas of productivity, participation, and compensation, among others. These essays will be valuable to the student or individual with only a basic knowledge of economics as well as to students of economic research and teachers of courses on women.

Studies in Economics, 8 cloth, $17.50; paper, $6.00

New from the National Bureau of Economic Research

THE ALLOCATION OF TIME AND GOODS OVER THE LIFE CYCLE

GILBERT R. GHEZ and GARY S. BECKER

This book is part of a growing literature on the economics of the household that views family formation, dissolution, acquisition of skills, and the use of resources as amenable to economic analysis. The authors develop a theoretical model which has wide applicability and strong explanatory power.

Human Behavior and Social Institutions, 6 $10.00

AN ECONOMETRIC ANALYSIS OF THE URBAN HOUSING MARKET

MAHLON R. STRASZHEIM

To provide an understanding of the economics of the housing market and its role in metropolitan development, the author applies econometric procedures to better assess the huge variations that exist in the characteristics and prices of available housing. A model is developed to explain the spatial variation in prices and incomes of occupants of the existing housing stock by neighborhood.

Urban and Regional Studies, 2 $12.50

COLUMBIA UNIVERSITY PRESS

Address for orders: 136 South Broadway, Irvington, New York 10533

ORGANIZATION FOR ECONOMIC COOPERATION AND DEVELOPMENT

Headquarters: 2, rue André-Pascal, 75775 Paris Cedex 16, France

OECD

economic growth
development aid
trade expansion

Periodical Aids for Reference, Teaching, and Planning
Available in English and French

	Subscription Rates *Surface*	*Air*
ECONOMIC OUTLOOK. A periodic assessment of economic trends and prospects in the main trading nations, published each July and December, $5 each. *"Occasional Studies,"* which treat selected topical problems in some depth, are included; or may be obtained separately ($2.50 each). Combined annual subscription:	$ 8.50	$ 10.50
OECD FINANCIAL STATISTICS. A collection of data on financial markets in 15 European countries, the United States, Canada and Japan. Of particular interest to investment trusts and other financial institutions. Issued yearly, with periodic supplements.	37.50	65.50
MAIN ECONOMIC INDICATORS. A unique tabulation of international indicators of economic activity. Graphics emphasize important changes in key indicators, and quarterly supplements provide additional data on consumer prices and industrial production. Monthly. Single copy, $2.50.	27.50	51.50
ECONOMIC SURVEYS. Annual review of each OECD country. Individual surveys, $1.75 each. Entire series:	22.50	41.50
PROVISIONAL OIL STATISTICS. Crude oil, gas, fuels. Output, trade and bunkers, by quarters. 4 issues yearly.	4.50	6.50
GUIDE TO LEGISLATION ON RESTRICTIVE BUSINESS PRACTICES IN EUROPE, NORTH AMERICA AND JAPAN. Includes a selection of relevant court and administrative decisions, up-dated by periodic supplements. Supplements, $10 per series, $16.00 by air. Basic volume:	50.00	
FOREIGN TRADE STATISTICS		
SERIES A. Overall trade by countries. Now monthly, plus half-yearly supplements.	30.00	45.00
SERIES B. Commodity trade, analysed by main trading areas. 6 volumes quarterly.	30.00	79.00
SERIES C. Detailed analysis by products. Imports and exports. Semi-annually.	35.00	107.00
LIAISON BULLETIN. Summarizes research under way in research and training institutes throughout the world. 4 issues yearly. Directories for Asia, Africa, Latin America and Europe available separately ($2.00 each).	6.50	14.00
NUCLEAR LAW BULLETIN. Legislative and regulatory activities, case law and administrative decisions, international organizations and agreements. 2 issues yearly, plus supplements.	4.50	7.30
PULP AND PAPER. Quarterly data on stocks of paper woodpulp, the production and trade of pulp and paper and board, the shipments of market pulp. Each issue, $3. Annually:	$9.00	
RESEARCH ON TRANSPORT ECONOMICS. Six-monthly information bulletin containing a subject index and an index of organizations mentioned. 2 issues yearly, May and November. (Sold only by annual subscription)	$28.00	
OECD OBSERVER. Illustrated bi-monthly magazine selectively covering the full range of OECD activities, and including announcements of new OECD publications. Sample copy on request. 6 issues yearly.	4.50	7.50

Rates subject to change.

MEMBER COUNTRIES

Australia
Austria
Belgium
Canada
Denmark
Finland
France
Germany
Greece
Iceland
Ireland
Italy
Japan
Luxembourg
Netherlands
New Zealand
Norway
Portugal
Spain
Sweden
Switzerland
Turkey
United Kingdom
United States

SPECIAL STATUS Yugoslavia

OECD also publishes about 100 monographs yearly.

A Blanket Order Plan is available.

Send orders and requests for information to:
OECD PUBLICATIONS CENTER
SUITE 1207E 1750 PENNSYLVANIA AVE., N.W.
WASHINGTON, D.C. 20006 • (202) 298-8755

The American Economic Review

CALCUTTA UNIVERSITY
CENTRAL LIBRARY

ARTICLES

The 1975 Report of the President's Council of Economic Advisers:

H. O. CARTER	Food and Agriculture
W. POOLE	Long on Analysis, Short on Policy
J. NIEHANS	Interest and Credit in General Equilibrium with Transactions Costs
K. LANCASTER	Socially Optimal Product Differentiation
T. W. EPPS	Security Price Changes and Transaction Volumes: Theory and Evidence
M. PAGLIN	The Measurement and Trend of Inequality: A Basic Revision
P. A. MEYER AND J. A. NERI	A Keynes-Friedman Money Demand Function
I. FRIEND AND C. LIEBERMAN	Short-Run Asset Effects on Household Saving and Consumption: The Cross-Section Evidence
A. G. SCHWEINBERGER	Pure Traded Intermediate Products and The Heckscher-Ohlin Theorem
W. L. SPRINGER	Did The 1968 Surcharge Really Work?
J. A. ORDOVER AND E. S. PHELPS	Linear Taxation of Wealth and Wages for Intragenerational Lifetime Justice: Some Steady-State Cases
J. A. FRENKEL AND C. A. RODRIGUEZ	Portfolio Equilibrium and the Balance of Payments: A Monetary Approach

SHORTER PAPERS: R. E. Wong; M. J. Beckmann; M. B. McElroy and J. C. Poindexter; K. B. Bhatia; E. J. Mishan; M. E. Burns; G. E. Mumy and S. H. Hanke; P. R. P. Coelho; A. Gifford, Jr. and C. C. Stone; P. Greenwood, C. Ingene, and J. Horsfield; J. R. Marchand and K. P. Russell; P. Bohm; R. Schmalensee; A. Daniere; L. C. Thurow; E. G. Davis and J. E. Olson; D. Weber; L. H. Meyer and W. R. Hart; P. L. Cheng.

SEPTEMBER 1975

THE AMERICAN ECONOMIC ASSOCIATION

Founded in 1885

- Published at George Banta Co., Inc., Menasha, Wisconsin.

- The American Economic Review, including four quarterly numbers, the *Proceedings* of the annual meetings, the *Directory,* and Supplements, is published by the American Economic Association and is sent to all members five times a year, in March, May, June, September, and December.

- 1975 membership dues of the Association are $23.00 a year, which includes a year's subscription to both the *American Economic Review* and the *Journal of Economic Literature.* Junior membership at $11.50 per year is available to registered students for a maximum of three years. 1976 membership rates will be available after October 1, 1975. Subscriptions (libraries, institutions, or firms) are $34.50 a year for 1975, and only subscriptions to both publications will be accepted. Effective January 1, 1976, the annual rate is $37.50; foreign postage add: PUAS, $2.10; all other countries and Canada, $2.50. Please write the Secretary's office, Nashville, Tennessee, for single copy prices.

- Correspondence relating to the *Papers and Proceedings,* the *Directory,* advertising, permission to quote, business matters, subscriptions, membership and changes of address should be sent to the Secretary, Rendigs Fels, 1313 21st Avenue South, Nashville, Tennessee 37212. Change of address must reach the Secretary at least six (6) weeks prior to the month of publication. The Association's publications are mailed second class.

- Second-class postage paid at Nashville, Tennessee and at additional mailing offices. Printed in U.S.A.

Officers

President

ROBERT AARON GORDON
University of California, Berkeley

President-Elect

FRANCO MODIGLIANI
Massachusetts Institute of Technology

Vice-Presidents

MARTIN BRONFENBRENNER
Duke University
CHARLES L. SCHULTZE
University of Maryland

Secretary-Treasurer and Editor of the Proceedings

RENDIGS FELS
Vanderbilt University

Assistant Secretary-Treasurer

C. ELTON HINSHAW
Vanderbilt University

Managing Editor of The American Economic Review

GEORGE H. BORTS
Brown University

Managing Editor of The Journal of Economic Literature

MARK PERLMAN
University of Pittsburgh

Executive Committee

Elected Members of the Executive Committee

IRMA ADELMAN
University of Maryland
ROBERT HEILBRONER
New School for Social Research
ANDREW F. BRIMMER
Harvard University
PAUL M. SWEEZY
Monthly Review
CAROLYN SHAW BELL
Wellesley College
BURTON A. WEISBROD
University of Wisconsin, Madison

Ex Officio Members

KENNETH J. ARROW
Harvard University
WALTER W. HELLER
University of Minnesota

THE AMERICAN ECONOMIC REVIEW

GEORGE H. BORTS
Managing Editor

WILMA ST. JOHN
Assistant Editor

Board of Editors

BELA BALASSA
EUGENE F. FAMA
MARTIN S. FELDSTEIN
ANN F. FRIEDLAENDER
STEPHEN M. GOLDFELD
ROBERT J. GORDON
ROBERT E. HALL
BENT HANSEN
JAMES R. MELVIN
WILLIAM D. NORDHAUS
LEONARD A. RAPPING
STEPHEN RESNICK
ANNA J. SCHWARTZ
JEROME STEIN
JOSEPH E. STIGLITZ
S. C. TSIANG
FINIS WELCH
MARINA V. N. WHITMAN

• Manuscripts and editorial correspondence relating to the regular quarterly issue of this REVIEW should be addressed to George H. Borts, Managing Editor of THE AMERICAN ECONOMIC REVIEW, Brown University, Providence, R.I. 02912. Manuscripts should be submitted in duplicate and in acceptable form and should be no longer than 50 pages of double-spaced typescript. A submission fee must accompany each manuscript: $15 for members; $30 for nonmembers. *Style Instructions* for guidance in preparing manuscripts will be provided upon request to the editor.

• No responsibility for the views expressed by authors in this REVIEW is assumed by the editors or the publishers, The American Economic Association.

• Copyright American Economic Association 1975.

September 1975

VOLUME LXV, NUMBER 4

CALCUTTA UNIVERSITY
CENTRAL LIBRARY

Articles

The 1975 Report of the President's Council of Economic Advisers: Food and Agriculture. *Harold O. Carter* 533

The 1975 Report of the President's Council of Economic Advisers: Long on Analysis, Short on Policy *William Poole* 539

Interest and Credit in General Equilibrium with Transactions Costs *Jürg Niehans* 548

Socially Optimal Product Differentiation *Kelvin Lancaster* 567

Security Price Changes and Transaction Volumes: Theory and Evidence *Thomas W. Epps* 586

The Measurement and Trend of Inequality: A Basic Revision *Morton Paglin* 598

A Keynes-Friedman Money Demand Function *Paul A. Meyer and John A. Neri* 610

Short-Run Asset Effects on Household Saving and Consumption: The Cross-Section Evidence *Irwin Friend and Charles Lieberman* 624

Pure Traded Intermediate Products and The Heckscher-Ohlin Theorem *Albert G. Schweinberger* 634

Did The 1968 Surcharge Really Work? *William L. Springer* 644

Linear Taxation of Wealth and Wages for Intragenerational Lifetime Justice: Some Steady-State Cases *J. A. Ordover and E. S. Phelps* 660

Portfolio Equilibrium and the Balance of Payments: A Monetary Approach *Jacob A. Frenkel and Carlos A. Rodriguez* 674

Shorter Papers

Profit Maximization and Alternative Theories: A Dynamic Reconciliation *Robert E. Wong* 689
The Limits to Growth in a Neoclassical World *Martin J. Beckmann* 695
Capital Gains and the Aggregate Consumption Function:
  Comment *Michael B. McElroy and J. C. Poindexter* 700
  Reply *Kul B. Bhatia* 704
The Concept and Measure of Consumer's Surplus:
  Comment *E. J. Mishan* 708
  Reply *Michael E. Burns* 710
Public Investment Criteria for Underpriced Public Products *Gene E. Mumy and Steve H. Hanke* 712
Externalities, Liability, Separability, and Resource Allocation:
  Comment *Philip R. P. Coelho* 721
  Comment *Adam Gifford, Jr. and Courtenay C. Stone* 724
  Comment *Peter Greenwood, Charles Ingene, and James Horsfield* 728
  Reply *James R. Marchand and Keith P. Russell* 730
Option Demand and Consumer's Surplus:
  Comment *Peter Bohm* 733
  Reply *Richard Schmalensee* 737
The Optimum Lifetime Distribution of Consumption Expenditures:
  Comment *Andre Daniere* 740
  Reply *Lester C. Thurow* 753
A Model of a Highway System Financed by a Trust Fund *E. G. Davis and J. E. Olson* 756
Optimal Advertising: A Note *David Weber* 759
On the Effects of Fiscal and Monetary Policy: Completing the Taxonomy *Laurence H. Meyer and William R. Hart* 762
Default Risk, Scale, and the Homemade Leverage Theorem: Note *Pao Lun Cheng* 768

Notes 774

The 1975 Report of the President's Council of Economic Advisers: Food and Agriculture

By HAROLD O. CARTER*

While *U.S.* crop production suffered the largest decrease in forty years, space devoted to a separate chapter "Food and Agriculture" (pp. 160–86) in this year's Economic Report reached record proportions. Since the inception of the Report in 1947, agricultural matters generally have been covered in subsections of chapters dealing with broader economic issues—and in about one-fifth the space. However, I resist the temptations to read greater significance into this happening than it probably warrants, having some knowledge and experience of the machinations and procedures for preparation of the Economic Report.

The focus of the food and agriculture chapter is largely on market-oriented agriculture and its interdependence with the world economy. The first section is a review of 1974 food prices, and it summarizes some of the favorable and adverse agricultural developments (pp. 160–65). The second and most lengthy section deals with long-term changes in agriculture (pp. 165–77). This is followed with a section on policy challenges and options (pp. 177–84). The chapter is concluded with a short section articulating guidelines for domestic farm policy (pp. 184–86).

I. Developments in 1974

Poor crops, substantial increases in marketing margins, and spiraling sugar prices were sighted as contributing to retail food prices that advanced at 13.4 percent annual rate between June and December 1974. For the entire year food prices rose 12.2 percent, second only to the record 20.1 percent in 1973.[1] Perhaps with discretion being the better part of valor, the Council refrained from a prediction for 1975.

Reverberations from the poor American crops were felt not only in our economy but on world markets. Indeed, our capacity to provide food aid to developing countries and even supply commercial export markets was severely tested. The Report stated: "The U.S. crop shortfall placed two new strains on the capacity to supply food aid. It first raised the opportunity costs of any given quantity of food aid, since any incremental exports would only aggrevate the adjustments required in the United States. It also raised the budgetary costs of any given volume of food aid during a period of concerted effort to hold down Federal expenditures" (p. 165). Most surely, the opportunity costs associated with food aid or even commercial exports to consumer and producer groups are different when supply is short as compared to past periods when we were disposing of burdensome surpluses.

II. Long-Term Changes in Agriculture

The Report states that since the early 1960's both the economic environment and the conditions of *U.S.* agriculture have

---

* Professor and chairman, department of agricultural economics, University of California, Davis.

[1] December to December percentage changes. See the Economic Report, p. 304.

undergone substantial change, "Excess capacity has declined, crop reserves have been drawn down, the world agricultural situation seems to have worsened, and agricultural products again appear to be subject to the unstable price conditions of an earlier era" (p. 165). While indeed some of the arguments presented in support of each of these developments is persuasive, the real question is the extent of the changes and whether they are transitory or represent permanent shifts.

Excess capacity declines in the post-World War II era are characterized by four developments according to the Report: (1) an apparent decline in the growth rate of agricultural productivity; (2) a shift from an excess supply of labor in agriculture toward a balance with demand; (3) a largely illusory acreage reserve; and (4) an expanded demand for *U.S.* output.

Consider first the apparent slowdown in growth of agricultural productivity. Total factor productivity (farm output per unit of total input) as measured by the U.S. Department of Agriculture (USDA) increased over each of the last three decades. During the 1950's it increased 27 percent compared to only 11 percent in the 1960's. The Report summarizes: " . . . The extent to which the dramatic decline in productivity growth in the 1960's represents a real and enduring decline is not clear, but the answer is of critical importance to the future trends of U.S. and world food supply" (p. 167). Serious questions have been raised over methods of measurement of agricultural inputs and outputs. A recent report of the National Academy of Sciences on "Agricultural Production Efficiency" cautioned considerable care in interpreting the trends in farm productivity or efficiency. Substantial revisions were apparently made by the USDA between 1971 and 1972. These changes operated to largely eliminate the apparent period of stagnation in efficiency exhibited in reports issued in 1971 and 1972. Admittedly, difficult conceptual and data problems exist in productivity measurement. The National Academy report notes several of these difficulties, including failure to reflect changes in quality of inputs. The case of labor is one example. No account has been taken of the substantial increases in the level of education of the farm labor force that has occurred over the past decade, or the changes in age and sex distribution. In conclusion, the National Academy report states "The available measures of inputs and outputs for U.S. agriculture are presently inadequate to provide evidence on the changes in the efficiency of agricultural production" (p. 38).

Excess labor in agriculture and income disparities between the farm and nonfarm sector according to the Report " . . . appear to be nearing an end" (p. 167). Support for this conclusion is drawn from farm employment and farm population data, both of which show a much slower rate of decline in the 1970's compared to the 1960's and 1950's. However, the Report contained some caution about overstating farm population which includes people residing on units officially defined as "farms." Changing life styles that make rural residences more attractive —not to mention changing census definitions of farms—may well be distorting our "true" farm count. Comparison of family farm and nonfarm incomes between 1970 and 1973 show a narrowing differential. "In 1970 the median income of farm families was about $3,700 less than that of nonfarm families; by 1973 the differential had been reduced to about $2,100" (p. 168). Extrapolations of such trends may overstate the closeness to equilibrium because 1973 was an unusually good year for agriculture. Moreover, another important dimension for shaping future farm policy is the distribution of income, within and between regions, which is not mentioned in the Report.

The agricultural reserve capacity attributed to land being withheld for production by government programs was generally assumed to be 60 million acres or about one-sixth of the nation's cropland. This has been referred to as illusory land reserve or "phantom acres" because they were not all returned to production after the federal acreage set-aside program was relaxed in 1973. In 1973 only 28 million additional acres came back into production. In 1974 farmers reported intentions to plant 17 million acres more but for various reasons including bad weather they expanded by only 9 million acres. Because of crop failures, cropland harvested was only 4 million acres more than in 1973. Thus, the Report points out that between 1972 and 1974 crop acreage rose by only 37 (28+9) million acres, suggesting less excess capacity than generally believed. The USDA's Economic Research Services indicate that crop prospects for 1975 suggest another 12 to 16 million acres coming into production, provided decent weather and prices prevail. The question of how much additional land could be brought into production generally brings forth a wide range of answers. To some, potential arable land refers to productive capacity or ultimate limits of land development without regard to cost or benefits. To others it is some unspecified point on the land development schedule, with assumed (but usually not clearly stated) levels of technology and prices. A USDA report provides estimates of how much *U.S.* land is suitable for crop cultivation. Of the total amount of the top three soil classes (I, II, and III), 365 million acres (58 percent) were being used as cropland and 265 million acres (42 percent) were in other uses (mainly forest and grass), but were considered suitable for regular cultivation. The 265 million acres of potential cropland was stratified according to feasibility for reclamation. They estimated about 152 million acres have high or medium potential for conversion to cropland.[2] While recognizing that agriculture is a land based industry, it is equally important to note that food output increases have come about from both intensive and extensive land use. The United States and most other developed countries have relied in modern times more on nonland inputs than on inputs of land to increase output. Thus, the concern expressed in the Report from a possible overestimate of excess land capacity because 23 million acres have been slow to reappear seems somewhat overstated.

The effects of increased demand for agricultural products as contributing to the reduction in excess agricultural capacity is discussed in the Report at some length. The volume of exports averaged 39 percent higher in the 1972–74 period than in the previous three years (fiscal basis). Again the real question is how much of the increase is temporary—due to poor world crops, two devaluations, and various market aberrations—and how much is of a more permanent nature and will continue as part of an upward trend? Even though agricultural exports for the fiscal year ending in June 1975 are forecast to reach a new record high of $22 billion, the volume of items shipped will likely be down by as much as 10 percent.

No mention is made of how much domestic demand for agricultural products will be affected by increased energy prices to the extent that those prices affect economic growth rates, general price levels, and income distribution. Some of our biggest commercial markets have scaled down orders as growth rates have slowed and declined in some cases. Japan and the United Kingdom suspended beef imports for short time periods. Other buyers have attempted to reduce trade deficits through cancellations, deferrals, and resales of food commodities. The effects from increased energy prices on those markets that exist in developing countries is more serious. A USDA report (1974) estimated that imports of petroleum, food, and fertilizer cost the developing countries $15 billion more in 1974 than in 1973. Of this total, oil import costs are expected to account for $10 billion, an amount roughly equal to developing nations' increased commodity exports during the past two years.

---

[2] This acreage was further divided into short-run and long-run. Short-run acreage (about 63 percent of the total) could be converted to crop uses virtually from one year to the next and long-run acreage would require construction of works of improvement and this takes a longer period.

The second subsection under "Long-Term Changes in Agriculture" is concerned with "A Changing World Agriculture" (pp. 170–72). The Report summarizes succinctly the world food-population balance controversy. While much has been written on this topic pre- and post-World Food Conference, the Report takes a middle-of-the-road position in terms of prospects for feeding an expanding world population. Three troublesome developments in world agriculture are noted: (1) a decline in the growth rate of world agricultural production; (2) a reduction in the supply of new land that can be brought into production, at least at supply prices of the past; and (3) the increased dependence on the United States as a supplier of agricultural products. The latter development is particularly alarming. "As recently as the late the late 1930's, North Africa, the Middle East, and Asia were net exporters of grains. Now these regions are consistently net importers. Similar trends elsewhere have made the United States the dominant exporter of grains, responsible for more than 50 percent of the total" (p. 172).

The Report (pp. 173–77) cites a number of factors to suggest that the *U.S.* food and agriculture sector has now entered a period of greater price instability: (1) large stocks held previous to 1972 provided a stabilizing influence on the market; (2) land reserve held in the 1950's and 1960's provided a further buffer offsetting changing conditions of supply and demand; and (3) there are expectations that an increasing share of *U.S.* output will go into export markets and thereby expose the agricultural sector to the vagaries of the world market. Fluctuations in world markets stemming from domestic policies of the European Community and Japan, and *USSR*'s growing involvements in world trade have exacerbated the effects on *U.S.* markets according to the Report. The Report notes that measures of variability for four important commodities (wheat, corn, beef cattle, and hogs) declined in almost every case comparing the 1910–49 period with 1950–71. For reasons not entirely apparent, the Report states that: "Without a significant rebuilding of stocks, more price instability should be expected than during the 1950's and 1960's, but it seems unlikely that year-to-year changes will be as large as in the earlier era" (p. 177). J. P. Houck's study on agricultural regulations and stabilization noted that individual prices of wheat, corn, soybeans, and cotton became more variable in the 1960's compared to the 1950's. He attributed this change to "market oriented" commodity programs and the use of direct income supplements to support farm income for producers of the key commodities. In any case, it would seem that we have too few observations in a period of conflicting events to generalize future trends with much confidence.

III. Policy Challenges and Options

The challenges are summarized under Agricultural Development, the Instability Problem, and Food Assistance. Under

Agricultural Development, the Rome World Food Conference is the lead item, but no assessment is given as to what it accomplished or what positive benefits can be expected in its aftermath. Certainly the dire predictions that were made by many notables at the opening last November that millions of people would die of starvation before mid-1975 have proven to be exaggerated. In fact, we have seen several months of declining food grain prices. This is not to imply complacency about a most severe food-population problem, but rhetoric won't solve it. The Report correctly focuses on the importance of developing and disseminating new production technology that will have a high rate of social return in the developing countries. Like the Rome Food Conference, the Report takes the less controversial course and avoids issues of population growth.

Grain reserves are discussed in the Report as one of the ways to alleviate the instability problem in agriculture (p. 180). There is some doubt expressed even to its need for that purpose if success is achieved in getting (1) improved worldwide information and analysis, (2) coordination with other countries of domestic agricultural policies, and (3) trade liberalization. The Report concludes that there are two reasons why the United States should build grain stocks above their current low levels. First, contingency reserves could furnish some price protection to *U.S.* consumers and producers (especially livestock producers). Second, it would guarantee our foreign customers steady access to supplies. Another reason not discussed would be to guarantee supplies of concessionary food aid to meet short-term emergencies in developing countries.

IV. Guidelines for Domestic Farm Policy

The principle enunciated in this section largely repeats what had been included in various Reports since 1970. That is to move toward greater market orientation with price supports below market clearing levels, limited production controls, if any, and direct cash payments, if needed. Yet it must be recognized that adherence to these principles is politically easier in periods of high agricultural prices and short supplies. President Ford vetoed recently a farm bill that would substantially have raised the target prices of the main commodities. While the veto was sustained this time, still lower prices would provoke another attempt with a better chance of success. Thus, until the question is answered of whether the economic environment in which the agricultural sector operates has substantially changed, it seems premature to "bury" land diversion, government payments, export subsidies, import restrictions and even mandatory production controls, no matter how appealing that option may be.

V. Concluding Comments

The chapter as a whole is well done and comprehensively summarizes most of the major issues and development in a "market oriented" agriculture in a world economy. It tries a bit too hard to establish the fact that "American agriculture finds itself in the mid-1970's at a watershed" (p. 165). From the title "Food and Agriculture" one might have expected more emphasis on food policy and less on farm policy per se. Some further recognition of the effects of agricultural and trade programs on consumer food prices as well as the viability of the agricultural sector would have been in order. The organization of the chapter leads to some repetition. For example, price instability is discussed as major subsections under two main sections. Similarly, agricultural development issues are treated in several sections with some overlap. Interestingly, only passing mention is made of the impact of increased energy prices on agriculture. Yet, President Ford

The 1975 Report of the President's Council of Economic Advisers: Long on Analysis, Short on Policy

By WILLIAM POOLE*

The Annual Report of the Council of Economic Advisers (*CEA*), which appears every year in late January or early February, is in many respects a peculiar political document. The function of the Report is in part to present the president's economic program, and in part to present the views of the *CEA* and its staff, a group of economists working in government for a few years and typically on leave from university appointments. There seems to be a tacit understanding that it is perfectly acceptable for the *CEA* to publish almost any academic views, irrespective of their political implications, and that what really counts are the explicit policy proposals.

This schizophrenic separation of the actual policy recommendations from the policy recommendations clearly implied by the general analysis is particularly evident in the 1975 Report. That the policy recommendations fall so far short of the analysis is probably due to a new president's difficulties in quickly putting together an economic program and to the fact that the policy implications of the *CEA*'s analysis are further removed than usual from what seems to be politically feasible. It is certainly not difficult to sympathize with the *CEA*, especially now that current political realities appear to leave it with relatively little real power. Having said all this, however, it is still the function of the economist-critic to discuss the economics and not the politics.

Besides the schizophrenia referred to above, two other features of the 1975 Report are noteworthy. One is the extent to which micro and macro issues are intertwined, and the second is the extent to which short-run policy prescriptions are dictated by long-run considerations.

In keeping with these introductory comments, the discussion below is divided into two major sections, one on the *CEA*'s macro analysis, and one on its policy recommendations. At the end are a few concluding comments.

I. Macro Analysis

The *CEA*'s macro analysis is contained in separate chapters on inflation and unemployment, and in bits and pieces in the chapters on the 1975 outlook, the 1974 review, and the appendix on the Economic Stabilization Act. By pulling together the pieces of analysis scattered throughout the Report, a coherent view of *CEA* attitudes on inflation and unemployment can be obtained.

A. *Inflation*

The *CEA* clearly accepts the monetary explanation for long-run inflation: "Rapid and sustained inflation requires a continual inflationary increase of the supply of money" (p. 131). This argument is repeated several times in different ways. Moreover, the argument that budget deficits create long-run inflation is explicitly rejected: "Yet price inflation caused by deficits with unchanging money supply would be of limited significance" (p. 131). Supporting the monetary explanation is a table (p. 133) showing inflation rates and money growth rates for 1965–70 and 1970–74 for the United States and five other developed countries.

---

* Brown University.

A monetary explanation for inflation would appear to place the responsibility in the lap of the monetary authority, and the *CEA* does not shrink from this conclusion. "Considering that our monetary authority, the Federal Reserve System, creates the quantity of reserves which is a basic determinant of how much money can be created in addition to hand-to-hand currency, one is led to ask why the regulation of the money supply has not prevented these undesirable price trends" (p. 133). Two paragraphs are then devoted to the argument that the Fed is not able to control money growth very precisely, but this argument is rejected.

> Nevertheless, the long-run growth in monetary aggregates is determined largely by the rate at which the monetary authority injects "high-powered money," . . . into the system. . . .
>
> . . . .
>
> The question therefore remains why the Federal Reserve System did not prevent this sustained period of steepening inflation. [p. 134]

Two reasons are given to answer "the question that remains." One is that the Fed financed part of the budget deficit and borrowings of federal credit agencies with monetary expansion . . . "Partly to avoid a tightening of the market to other borrowers . . . " (p. 135). The second reason for the high rates of monetary expansion . . . "was the effort of policy makers to play safe against recessions . . . " (p. 135).

Following this discussion, and complementing the analysis of the Phillips curve in the preceding chapter on unemployment, is a section on "The Unstable Tradeoff" (pp. 136–41). The thrust of this argument is that there is no long-run tradeoff between inflation and unemployment, although there may be a short-run tradeoff due to lags in price adjustments. This discussion, along with comments on the deficiencies of wage and price controls, has as its basic message that we had better take our medicine now: " . . . there is always a choice between accepting the difficulties of adjustments 'now' or moving toward even greater future difficulties" (p. 137).

The *CEA* analysis of long-run inflation is correct; prices could not have risen so rapidly without the monetary expansion, and the monetary expansion was under the control of the Federal Reserve. The validity of the analysis of the reasons for the excessive monetary growth is less clear; it is always more difficult to discuss motives than facts. It might, however, have been useful for the *CEA* to sidestep the question of what motives were in fact operative by concentrating on the effectiveness of the actual policies. The argument that the long-run Phillips curve is vertical makes clear that excessive monetary expansion is not in fact an effective way to "play safe against recession." Similarly, the Council mentions that high monetary expansion is not an effective way " . . . to avoid a tightening of the [credit] market to [nonfederal] borrowers . . . " (p. 135), but the point deserves a more extensive treatment, perhaps along the lines of the Phillips curve analysis, to show that as inflation develops it is only a matter of months before interest rates end up higher than they would have without the extra monetary expansion. Since the interest rate peak is ultimately higher, the damage to interest-sensitive markets such as the mortgage market is actually exacerbated rather than eased by Federal Reserve efforts to hold interest rates down.

Another point of interest in the *CEA*'s analysis of inflation is its clear discussion of the impact of inflation on federal tax revenues. In the review of 1974 (ch. 2), it is noted that " . . . from the first quarter of 1973 to the third quarter of 1974 the full-employment surplus rose by $35 to to $36 billion" (p. 63). Most of this increase is attributed to the effects of inflation, and at the end of the chapter on inflation there is a fairly detailed analysis of the features of the tax structure that give rise to the high elasticity of revenues with respect to inflation.

While most of the analysis concerns inflation in the long run, there are a few comments concerning the causes of inflation in the short run. The thrust of the argument is that special factors may change the timing but not the ultimate extent of price rises caused by excessive money growth. Among the special factors mentioned as having some importance in 1974 were the end of wage-price controls and the end of fixed exchange rates. These factors are mentioned as specific examples of a more general lag phenomenon. Although the noneconomist will probably not find the exposition very clear, the *CEA*'s position is that rising unemployment can be expected to be associated with a price deceleration, so that the lags are to be interpreted in terms of changes in the rate of inflation rather than in terms of the price level itself. Moreover, the *CEA* argues that the price decelerations associated with recessions have become steadily smaller over the postwar years, the implication being that progress in reducing inflation will be slow and painful.

The inflationary contribution of the Organization of Petroleum Exporting Countries (*OPEC*) is mentioned, but underplayed. The impact of the *OPEC* cartel on the aggregate price level is said to be "substantial" but no effort is made to pin much of the blame for recent inflation on *OPEC*.

B. *Unemployment*

Chapter 3, "Unemployment," is substantially longer than chapter 4, "Inflation During the Last Decade," and to my tastes less interesting. While the Council says that "This chapter is intended to be a guide to the formulation of constructive policies toward unemployment over the long run" (p. 86), the guidance offered is almost entirely in terms of explaining the meaning and interpretation of the data rather than in terms of an analysis of the causal factors involved and of the policies required. The measurement of unemployment in the United States is explained, and *U.S.* figures compared to those in other developed countries. Unemployment is categorized into frictional, structural, seasonal, and cyclical components. The concepts of duration per spell and number of spells per worker are explained. Unemployment in age, sex, student/nonstudent, veteran/nonveteran, educational attainment, and race categories is described. Finally, there is a brief description of the major income maintenance programs.

While most of the chapter on unemployment involves a description of the data, two analytical themes are prominent. The first is that there is no long-run tradeoff between unemployment and inflation, even after adjusting for changes in the demographic composition of the labor force. The second theme—one that appears in scattered comments rather than in an integrated argument—is that many government programs add to the level of unemployment. Mentioned are minimum wages (p. 107), state equal pay laws (p. 113), and unemployment benefits (pp. 121, 123–24).

The big disappointment in the chapter on unemployment is the utter lack of any discussion concerning the causes of the current recession; the *CEA*'s silence on this issue is in sharp contrast to its analysis of the causes of long-run inflation. In the light of the popular view that recessions are the result of inherent instability in the private sector, it would have been very useful for the *CEA* to have emphasized the fact that the current level of unemployment is the direct result of low monetary growth in 1974 following high monetary growth in 1972–73 and a huge rise in the full-employment budget surplus in late 1973 and 1974. Once the policy mistakes of 1972–73 had been made, a recession of some sort was perhaps inevitable, though the 1974 policies exacerbating the contraction were not. In any event, whether one emphasizes the policy mistakes of 1972–73, or of 1974, it is clear that recessions of the current magnitude are made in Washington, not New York, Detroit, Chicago, or Los Angeles.

II. Macro Policy Proposals

If the *CEA* Report had been written as I have written about it—a section on analysis and a section on policy—the reader would have come to the policy section expecting to discover a substantial list of policy programs. After finishing the Report what he discovers instead is an extraordinarily meager policy program.

The stabilization program—along with the energy proposals—is presented and defended in one short chapter (ch. 1, "Economic Policy and Outlook") of 15 pages. There is no point in presenting a detailed examination of the proposals; they were generally recognized as being obsolete six weeks after the Report was issued. In outline, the policy proposals involved: a) an excise tax of $2 per barrel applied to both domestic and imported oil and an equivalent tax of $0.37 per thousand cubic feet on natural gas; b) reduction in individual and corporate income taxes just offsetting the new energy taxes; c) temporary tax cuts in the form of a one-year increase in the investment tax credit and a tax rebate for individuals of 12 percent of their 1974 tax liabilities, up to a limit of $1,000 per return.

It is interesting, and probably unfortunate, that the stabilization and energy policies are discussed together. The association of the two tends to mix up stabilization issues with micro issues, but this mixing was perhaps inevitable once the decision was made to deal with the micro problems by levying new taxes large enough to have macro implications.

The *CEA* did not expect much of a stabilization effect from this modest program. It was estimated that the program would cut the extent of the forecasted year-over-year decline in income in 1975 " . . . one-half of 1 percent to 1 percent in terms of real GNP" (p. 20). This stimulus, by Okun's Law, amounts to 1/6 to 1/3 of a percentage point of unemployment. Given the forecast that " . . . the unemployment rate should average about 8 percent for the year . . . " (p. 27), it is clear that the *CEA* felt unwilling or unable to recommend fiscal policies that promised to do very much.

The small fiscal program did not, it is clear, stem from an assumption that monetary policy would do the job. In contrast to the 1971 Report, where it was argued that "with the proposed budget and complementary monetary policy," a *GNP* in the neighborhood of $1065 billion was a "reasonable expectation" (p. 85), nowhere in the 1975 Report is there a clear statement that a monetary policy assumption is needed to generate the near-term economic forecast.[1] The contrast with the *CEA*'s explicit discussion of the role of monetary policy in long-run inflation is so great that a rather lengthy quotation is in order.

---

[1] The *CEA* does, however, use language suggesting that it may favor a more expansionary monetary policy. It says that "one way" of preventing federal deficits from crowding out private investment would be to increase the rate of money growth and that in a substantially unemployed economy ". . . an increase in monetary growth need not be inflationary in the short run . . ." (p. 25).

> Monetary policy faces great difficulties in the year ahead and will require careful and continuous evaluation by the Federal Reserve. The uncertainties that underlie the outlook for 1975 add to the importance of a flexible monetary policy. Monetary policy must be conducted so as to encourage near-term recovery in the economy and a resumption of sustainable economic growth. Toward this end, reasonable growth in money and credit will be required—growth which, one hopes, will encourage a freer flow of credit and lower interest rates in private credit markets. Whether more accommodating credit conditions will in fact develop depends importantly on the ease with which the enlarged Federal deficit is financed, and also on the progress that is achieved in moderating the Nation's rate of inflation as 1975 progresses. [p. 26]

This nonstatement is inconsistent with views expressed elsewhere in the Report. Not only is the analysis of long-run inflation quite explicit on the Fed's role, but in chapter 2, reviewing the events of 1974, the *CEA* says that, "Growth rates in monetary aggregates and in total spending tend to move in similar patterns . . . " (p. 66). In this same chapter the sharp decline in money growth in the second half of 1974 is noted; in addition, the *CEA* says that a " . . . slightly more expansive set of monetary actions. . . . [is] expected to show soon in the monetary aggregates" (p. 67).

Neither this Council nor any previous one has been willing to face up to the issue of what happens when monetary policy is inappropriate, however judged by the particular Council. In the light of the Federal Reserve's admitted power to generate boom or bust, the unwillingness to take on the Fed is peculiar. The standard *CEA* position must reflect the Fed's extremely strong political position and the expectation held by each Council that the Fed will follow sensible policies. But it seems inevitable that a time will come when a Council, or a President, or a Congress, will take on the Fed if the Fed's performance continues its past pattern.[2] Whatever the abstract merits of an independent monetary authority, independence is in fact doomed if money growth always turns out to be high during inflation and low during recessions.

In the 1975 Report the *CEA* is obviously not expecting a continuation of the low monetary growth rates of late 1974, but neither is the *CEA* calling for any real monetary stimulus. Indeed, the general tone of the Report is that policy instruments should not stray far from the settings consistent with long-run price stability. It is understood, and accepted, that such policy settings will produce a long period of above-normal unemployment. The reasons for the *CEA*'s acceptance of a long period of slack are not entirely clear, but they seem to involve a mixture of economic and political considerations. The two economic arguments against a more expansionary fiscal policy are that " . . . a too expansionary budget carries the risk of worsening the inflation . . . " (p. 24), and that deficit financing will crowd businesses out of the financial markets.

The first argument is certainly correct if what is meant is that more expansionary policies will make the rate of inflation higher than it otherwise would have been. The noneconomist, however, would probably interpret the language used as involving the argument that more expansionary policy runs the risk of making inflation higher than it was last year. With the unemployment rates projected by the

---

[2] In early 1974 Congress took a mild step in the direction of greater control over the Federal Reserve by passing House Concurrent Resolution 133 which requires periodic reports by the Federal Reserve to Congress on monetary policy plans.

CEA this argument would seem untenable. There is no effort in the Report to examine the issue of how fast inflation ought to be reduced.

The second argument made against a more expansionary fiscal policy—the crowding out argument—appears incorrect within the context of the conventional Keynesian model in a time of extensive unemployment. For a given investment demand function and a given rate of money growth, the effects on real *GNP* of a more expansionary fiscal policy will be partly but not completely offset by a decline in private investment. Expansionary fiscal policy raises both real income and nominal interest rates. Moreover, in the conventional model the rise in income would, through the investment accelerator, cause the investment demand function to shift out, further increasing both income and interest rates. This latter argument might be especially important if the slack in the economy had pushed many firms close to bankruptcy. Loss of creditworthiness is surely a much greater threat to financing new investment than is a small rise in interest rates.

Is there anything wrong with the conventional argument? The *CEA* itself accepts the conventional argument as applied to past recessions but not as applied to the current one, the difference being that financing demands are said to be greater in inflationary conditions. Perhaps what the *CEA* has in mind is a proposition that the money demand function is highly interest inelastic under inflationary conditions so that, with a given investment demand function, an expansionary fiscal policy produces a relatively large increase in interest rates and a relatively small increase in income. In this case, for firms already overburdened with debt, the impact of rising interest rates might outweigh the favorable effects of rising sales, thus forcing weaker firms into bankruptcy and generating a depressing wave of business and investor pessimism.

Another possibility is that investment demand is more interest elastic in inflationary than in noninflationary conditions as suggested by the argument that, "As the desired private refinancing is made more difficult by the deficit financing, businesses may abandon investment projects more readily than in the past, rather than risk further unbalancing their capital structure and increasing their credit market exposure" (p. 25). To the extent that this argument is correct, an expansionary fiscal policy will not raise interest rates very much—only a small increase will be required to crowd out private investment—nor will fiscal policy be of much value in reducing unemployment. However, if the deficit were smaller and interest rates lower, would firms be willing to "further unbalance their capital structure" in order to raise debt capital to finance new investment? The logic of the argument would seem to require that federal deficit financing crowd out private equity financing.

While I am not convinced by the argument that fiscal policy is less useful under inflationary conditions, I would add that the *CEA* is probably correct in not expecting much benefit from a more expansionary fiscal policy. The crowding-out argument has increasing force as the time horizon is extended. The initial stimulus from fiscal policy is offset, either entirely or mostly, later on. If weak private demand is likely to continue for five years—a not unreasonable projection—then fiscal policy is probably not the key to stabilization policy at this time. It may well be useful to have some fiscal stimulus now in order to limit the depth of the recession, but such stimulus, even if continued, will make little difference to the unemployment rate, and possibly even make the rate higher than it otherwise would have been, three years from now. Monetary policy could make more of a difference three years from now, but only at the cost of generating a higher inflation rate than otherwise would have occurred. For a given course of policy, the key to guessing what will happen is the speed with which inflation—and the expectation of inflation—declines. Great trouble can be caused by a 5 percentage point decline in the rate of inflation, whether from a starting point of 10 percent inflation, or of price stability. In recent years firms and households have committed themselves to a tremendous amount of debt bearing high nominal rates of interest; the *ex post* real rate of interest on this debt will be painfully high if the rate of inflation adjusts downward quickly.

What the *CEA* does not make clear is that the policy issue is not one of more or less, but of more now or more later. There is, of course, no law of "conservation of unemployment"; the total man-years of unemployment over the next ten years is not a constant and there may well be a positive, or negative, return to taking unemployment in one big lump now rather than gradually over a period of time. But an important element in the policy problem is that of trading off present against future and the *CEA* would have performed a public service by following its analysis of the lack of a long-run tradeoff between unemployment and inflation with an analysis of tradeoffs between present and future.

In addition to the arguments already discussed, another *CEA* reason for rejecting a more expansionary fiscal policy is almost certainly a deep concern over the rising share of *GNP* taken by government. This position is expressed most clearly in the President's Report where it is asserted that total government expenditures now amount to one-third of *GNP* and have been growing at a rate such that in twenty years government will take one-half of *GNP*. "Spending on this scale would require a substantial increase in the tax burden on the average American family. This could easily stifle the incentive and enterprise which is essential to continued improvements in productivity and in our standard of living" (p. 4).

The micro issue of the allocation of total output between the government and private sectors frequently intrudes into the macro issue of stabilization policy. Such an intrusion is not surprising—tax changes for stabilization purposes are almost invariably delayed by arguments about who should benefit or who should lose—but in this case the statement in the President's Report is seriously misleading. The share of total government purchases, in contrast to expenditures, in *GNP* in 1974 was 22 percent, up from 20 percent in 1960, and actually down from 23 percent in 1968. Moreover, the projections of government purchases contained in the 1970 Report (Tables 13 and 14, pp. 79, 80), converted to 1974 dollars were for state and local government purchases only $0.7 billion below the 1974 actual, and federal purchases $17.2 billion *above* the actual.

The real issue is the rapid growth of total government transfers to persons. In 1974 these transfers amounted to 10 percent of *GNP* and 12 percent of personal income, up from 5 percent and 7 percent, respectively, in 1965. Actual federal government transfers to persons in 1974 were $18.4 billion (19 percent) above the projections for 1974 made in the 1970 Report, and the projections had already included an allowance of $6.6 billion for the Family Assistance Plan of welfare reform.[3]

Transfers do not per se increase the net burden of taxes less transfers for the average family, but they may have significant disincentive effects. For those whose disposable income is raised the disincentive effect may be substantial since the income effect and the substitution effect generated by higher marginal tax rates both work in the same direction. As noted earlier, the *CEA* in its discussion of unemployment did mention the disincentive effects of some income maintenance programs.

---

[3] The 1970 projections for transfers were converted to 1974 dollars using the personal consumption deflator.

Income maintenance programs have grown so rapidly in recent years—a growth that is likely to be extended and accelerated by the current recession—that Congress will have to take up the issue of welfare reform once again. With government transfers amounting to 12 percent of personal income in the context of a poorly designed set of programs, the disincentive effects may well be large enough that welfare reform should be treated as having substantial macro implications—implications for aggregate real *GNP*—as well as the obviously important humanitarian and distributional implications. It is disappointing that the *CEA*, in the light of its concern over the problems of incentives and excessively rapid growth of government spending, did not produce a better analysis of the issues involved.

III. Concluding Comments

The *CEA* presents essentially no program for dealing with the current recession. There is no explicit analysis of how the current recession came about, and no analysis of the choice that must be made between the speed with which inflation is to be reduced and the speed with which unemployment is to be reduced. If the *CEA* rejects the notion of a short-run tradeoff under present conditions—an argument that, whether right or wrong, certainly could have been made—then it should have presented the details of a positive case for not following expansionary policies under current circumstances.

The *CEA* analysis clearly demonstrates the need for long-run reforms in the way macro policy decisions are made, but no program of reform is offered. One proposal for reform might have been the adoption of a presumptive norm for monetary expansion. For example, the norm might be 3–5 percent growth for M_1, or 4–6 percent for M_2, to be breached only upon an explicit policy decision that circumstances require a departure or that the long-run norm must be changed. It would also have been possible to present a lengthy list of reforms for improving the accuracy of monetary control.

Fiscal reforms are also urgently needed. As noted earlier, the welfare reform issue must be faced again. From the *CEA* analysis of the effects of inflation on government revenues, it is clear that indexing of the tax system would be desirable. Indeed, the *CEA* says that "... individuals and corporations alike are now exposed to substantial and haphazard tax-raising effects produced by inflation, even during a period of falling real incomes ... adjustments in tax laws will be called for during the gradual process of price deceleration toward which we are heading" (p. 145).

Yet in the very next paragraph the *CEA* rejects such adjustments on the grounds that larger federal deficits will increase strains in the credit markets and thereby add to pressures on the Fed for higher monetary expansion. But surely, if the *CEA* analysis is correct, an indexed tax system that stabilized the full-employment surplus and prevented unforeseen inflation from generating unforeseen and unintended changes in fiscal policy is consistent with a rapid reduction in actual deficits as the recovery proceeds.

The lack of interest in long-run reforms in macro policy does not, apparently, reflect a general unwillingness to propose long-run programs. The *CEA* does discuss long-run proposals in the areas of energy, government regulation of industry, and agriculture, with separate chapters being devoted to analysis and policy proposals in the latter two areas.

The unwillingness of the *CEA* to present a package of reforms in the macro area reflects two failures in macro policy analysis in the economics profession in general. The first is that the advocates of discretionary policy have generally refused to be pinned down on formulations of optimal policy reaction functions. Policy makers are too often encouraged to respond "flexibly" in an "appropriate" fashion to the current "needs" of the economy.

There is no hope of improving discretionary policy without the explicit specification of policy reaction functions that are to be followed until evidence accumulates that better policies are available. If the policy record is to be evaluated so that past mistakes can be avoided in the future, then the policy problem cannot be viewed as one of coping with events unique in time and place. The extent to which policy makers place their bets in accordance with the objective probabilities in inherently uncertain situations cannot be judged if each situation is viewed as a separate drawing from a separate probability distribution; the probabilities in a stochastic process can never be determined from single outcomes.

The second general failure of macro policy analysis—one that apparently troubled the *CEA* in its 1975 Report—is that of presenting the basic policy issues in such a way that they can be understood by the general public. Macro policy is determined not by economists but by the political process. The destabilizing nature of monetary and fiscal policy on average since the Employment Act is not the result of perfidity but of ignorance. Congress and the general public simply do not understand the consequences of monetary and fiscal actions. In numerous places in the 1975 Report the *CEA* seems to back away from the policy implications of its analysis out of a fear that sensible proposals might be bent in directions that will make us worse off than if we do nothing. How, for example, can advocacy of a smaller government in the long-run be consistent with advocacy of government deficits now? How could advocacy of rapid money growth now be consistent with advocacy of slow money growth later? It is to be hoped that future *CEA* Reports will meet these issues head-on, as some of the past Reports have on some issues, because surely the major gains in macro policy in coming years will come from applying what we already know rather than applying the latest findings appearing in the professional journals.

REFERENCE

U.S. Council of Economic Advisers, *Economic Report of the President,* Washington, Feb. 1970, 1971, 1975.

Interest and Credit in General Equilibrium with Transactions Costs

By JÜRG NIEHANS*

It has long been recognized that the key to a better understanding of the services of money in general equilibrium is a theory of market imperfections or "frictions" (see, for example, John Hicks). There has been a growing interest in monetary exchange models in which market imperfections appear in the form of transactions costs (see F. H. Hahn (1971b), Mordecai Kurz, Kevin Sontheimer, Ross Starr, Edi Karni). Progress has been made in analyzing the efficiency of equilibrium in such models (Hahn (1973), David Starrett). However, it has not so far been possible to free the analysis from arbitrary assumptions about terminal money stocks. There is also no explicit analysis of the demand for cash balances relative to commodity stocks, interest-bearing assets and debt, nor do we know the exact conditions which in a monetary economy with transactions costs correspond to the familiar marginal rules in neoclassical models. The problems so admirably identified, but imperfectly solved, by Jacob Marschak in his 1949 paper, continue to absorb the attention of monetary theorists.

An explanation of the transactions demand for money and bonds, based on transactions costs, was offered by William Baumol and James Tobin. Edgar Feige and Michael Parkin have recently generalized this approach by allowing for the repercussions of transactions costs and portfolio income on consumption. However, they still confine themselves to optimization by a representative individual at exogenously given prices and rates of return. It is not explained who holds the cash balances and the bonds which, at a given moment, the representative individual does not plan to hold. Correspondingly, there is no effort to explain commodity prices and interest rates by the interplay of demand and supply for commodities and assets. Feige and Parkin also give no explanation of private debt, and they continue to use the *ad hoc* assumption of a regular saw-tooth profile of the excess of income over consumption on which all of their results are crucially dependent. A general equilibrium counterpart to the Baumol-Tobin approach is lacking.

The present paper combines elements of the general equilibrium approach and of the Baumol-Tobin approach to solve problems left open by each of them. The result is a general equilibrium model of a monetary economy with transactions costs, explaining the stock demands for commodity inventories, cash balances and bonds, and the supply of private credit. In contrast to the Feige-Parkin model, it satisfies the requirement that all cash balances and bonds must be willingly held by somebody at all times, and prices and bond yields are endogenously determined. It requires no arbitrary wealth constraint, and the time profile of assets is derived from the economizing behavior of economic agents. In an earlier paper (1971), I have shown that with given endowments and tastes, different exchange arrangements ranging from bilateral barter to full monetization and neutral money can be made to emerge simply by varying the transactions and storage costs. Its most serious limitation was the neglect of credit and interest. It is the purpose of the present paper to fill this gap.

---

* Professor of political economy, The Johns Hopkins University.

Section I outlines the general features of the model, describing an exchange economy in infinite stationary motion where individuals, subject to time preference, maximize the utility of consumption by holding commodities, money, bonds, and debt. Section II formalizes the non-linear programming problem of an optimizing individual, subject to the multiple constraints typical for transactions cost models; it also presents the marginal conditions on consumption, exchange flows, and stocks. The conditions for market equilibrium are specified in Section III, and it is shown that they impose limits on the admissible range of asset yields. In the final section, the model is used to derive the implications of an "optimal" money supply rule for the demand for cash balances, welfare, and the demand and supply of credit.

I. The Model in Outline

A. *Assets*

The analysis relates to an exchange economy without production. Individuals are endowed with certain quantities of consumables. To focus attention on the relationship between commodities (as a group) and financial assets, all consumables will be lumped together into one commodity or good. There are three assets: commodity stocks, money, and bonds. Bonds have the form of a standardized perpetuity. Certain aggregate quantities of money and bonds are supplied by the government, which finances its interest payments by lump sum taxes. Bonds may also be supplied by private individuals selling and repurchasing their own consols. Debt thus appears as a negative stock of bonds. It is assumed that the economy is fully monetized, using token money as universal medium of exchange.[1] There are thus two markets, one for goods and one for bonds.

The net yield of each asset consists of its market return, the capital gain or loss, and time preference. The market return on commodity stocks is negative, consisting of a storage cost expressed as a physical shrinkage. Bonds have a uniform and constant coupon r, fixed by the government and payable in money; their market return i fluctuates inversely to their price according to $i=r/p^b$. There may also be a positive market return on cash balances, fixed by the government. Capital gains and losses are created by changes in commodity and bond prices. The analysis thus incorporates the speculative motive for asset holding familiar from Keynesian models and stressed by Marschak. However, assets with positive net yields are not compatible with equilibrium, because further accumulation always pays. This problem is often evaded by an arbitrary wealth constraint. In the present model it is solved by introducing time preference as a third, "psychological" component of asset yield. The total net yield of an asset can thus be nonpositive even though there may be positive market returns and capital gains.

Despite the possible differences in the yields of different assets, portfolios may be diversified due to transactions costs. An individual will prefer commodity stocks to money balances for short periods, because the saving in storage costs is not worth the transactions costs on the exchange of commodities. For sufficiently short periods, he will also prefer non-interest-bearing cash balances to bonds, because the positive return on bonds is outweighed by the transactions costs on bonds. In a pure exchange economy, all assets have the character of transactions assets, used to "bridge the gaps" between the given time profile of endowments and the preferred time profile of consumption. Debt is an alternative way to achieve the same purpose; it is profitable whenever the resulting savings in asset-holding costs at least pay for the additional interest and transactions costs.[2] Transactions costs are assumed to have the form of commodity inputs,[3] proportional to the quantity exchanged.[4]

---

[1] These assumptions make it impossible, of course, to analyze relative commodity prices, the differences between monetary exchange and barter, and the choice between different media of exchange. For a treatment of these problems the reader is referred to the author's 1971 paper.

B. *Time and Utility*

For an analysis involving stocks, the temporal structure of the model is of crucial importance.[5] For the present model, this structure can be described as one of *infinite stationary motion*. Let time be divided into 1 . . T . . H years, while each year consists of 1 . . t . . h days.[6] The model is dynamic in the sense that things are subject to change from day to day within each year. This provides the basis for the demand for stocks. The model is stationary in the sense that nothing is allowed to change from year to year. This makes it possible to construct a counterpart to the stationary equilibrium of neoclassical models. In particular, endowments are the same for a given calendar day of each year, tastes remain the same from year to year, and the individual will be constrained to choose a plan with identical consumption, exchange flows, and stocks for the same calendar day of different years. H is infinite and foresight is perfect. We can thus visualize all decisions as being made at the beginning of time with transactions costs becoming due when the deliveries are made. The two parts in each exchange are assumed to be simultaneous; a sale of goods with deferred payment is constructed as a cash sale combined with a bond purchase. The concept of infinite stationary motion is of decisive help in explaining the indirect utility of assets, including cash balances, without arbitrary assumptions about initial and terminal stocks.[7]

Individuals maximize utility over the infinite horizon.[8] Utility depends on consumption only; money, inventories, and bonds may have indirect utility by permitting a more desirable pattern of consumption. The form of the utility function must conform to the stationary motion framework. It is thus assumed that tastes remain the same from year to year, but may vary from day to day within a given year. Let $U=U(x_1 \,.\,.\, x_t \,.\,.\, x_h)$ be the annual utility function as seen at the beginning of the year. It is assumed to be twice differentiable and to have diminishing marginal rates of substitution and positive marginal utilities. Note that the function is not additive, thus allowing for daily variations in tastes and for intertemporal substitution and complementarity, which are presumably some of the determinants of the transactions demand for assets. Utility over the infinite horizon is assumed to be the discounted sum of annual utilities. Since the individual is constrained to plans which are stationary from year to year, maximization of infinite-horizon utility is equivalent to the maximization of annual utility. As a consequence, $U(\ \ )$ is all we have to pay attention to.

---

[2] In the absence of time preference and storage costs, nobody would go into debt, since asset holding would be costless. This is the basic reason why the Baumol-Tobin approach cannot explain the supply of debt.

[3] From a fundamentalist point of view, transactions costs may appear to be largely caused by uncertainty and the consequent search for information. While these elements are important in the world as we know it, recent contributions have shown that a meaningful general equilibrium model of monetary exchange can be constructed on the assumption of full certainty. For efforts to develop an explicit analysis of search, see the papers by Karl Brunner and Allan Meltzer, and by Joseph Ostroy.

[4] It is clear that scale economies in transactions costs, depending on transactions scale as well as on market scale, play an important role in actual exchange arrangements. While such scale economies have often been prominent in models of individual optimization (as, for example, the Baumol-Tobin and the Feige-Parkin models), they are, for well-known reasons, difficult to incorporate in a general equilibrium system (for a recent contribution see Walter P. Heller). The following analysis indicates that the basic aspects of a money and credit economy do not arise from scale economies.

[5] In Hahn's (1973) terminology, one of the requirements is that the economy allows for "not inessential sequences."

[6] Flows with index t are concentrated at the *end* of day t. The paper offers no analysis of the sequence of events during the trading period of a single day. On this point see E. C. H. Veendorp and the author's 1973 note.

[7] Hahn (1973) conjectured that the assumption of exogenously given terminal stocks can be abandoned in an economy with infinite horizon and overlapping generations (p. 456). Indeed, this seems to be one type of economy capable of infinite stationary motion.

[8] To prevent a frequent error, it may be noted at this point that utility maximization generally does not imply transaction cost minimization. Nor does the transactions cost approach to monetary equilibrium maintain that the use of money lowers aggregate transactions costs. Transactions costs are indeed minimized for a given volume of transactions (as in the author's 1961 paper), but this may be more than balanced by a higher volume of transactions.

Time preference is implicit in the utility function. In particular, the derivative $\partial U/\partial x_t$, to be called the *discounted marginal utility*, depends in part on the time span between the planning moment and day t. It will gradually increase as t comes nearer. To make the role of time preference explicit, we define $\partial V/\partial x_t=(1+\theta)^t(\partial U/\partial x_t)$, to be called *marginal utility* (without qualifying adjective). It is meant to represent marginal utility "purged" of time preference, and it is postulated to be invariant as we approach it.

C. *A Graphical Illustration of Portfolio Diversification*

The diversification of assets is illustrated, nonrigorously and heuristically, in Figure 1. It depicts the annual cycle of an individual's optimal portfolio, abstracting from daily price changes, interest on cash balances, and taxes. All of its elements are jointly determined; none are exogenously given.

Endowments accrue to the individual in the form of commodities. Unless the time profile of endowments, by coincidence, matches the time profile of preferences, the individual will want to divorce the pattern of consumption from the pattern of endowments by holding assets. Total

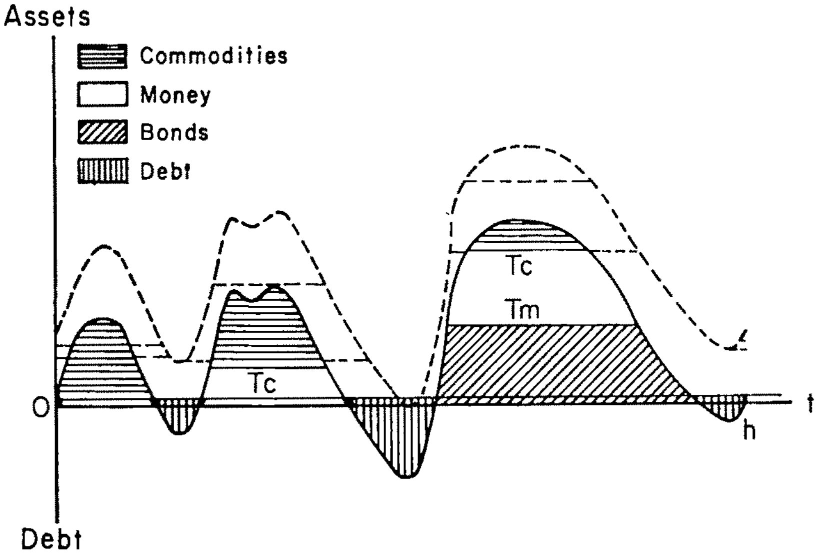


FIGURE 1

assets as a function of time are described by the solid curve. The stationarity postulate implies that this curve begins and ends the year at the same level. We assume for convenience that it begins the year at zero.[9] Total assets may, of course, be negative, indicating debt. The layers revealed in a vertical cross-section through the asset profile indicate what portions of total assets are held in the form of commodities, money, bonds, or debt, respectively, at a given moment. A horizontal section shows how long the marginal unit of an asset is held.

Commodity stocks will be held if the holding period is so short that the transaction costs for a return trip into cash balances would be higher than the storage cost. This determines a critical holding period for commodities, T_c. In the absence of transactions costs on goods, no commodity stocks would ever be held. Cash balances will be held if the holding period, though exceeding T_c, is still so short that the transactions costs for a return trip into the bond market would be higher than the market return on the bonds. This determines a critical holding period for money, T_m. In the absence of transactions costs on bonds, no cash balances would ever be held, nor will cash balances be held if $T_m < T_c$.

An individual goes into debt in order to shift his asset curve downward. People borrow to make it unnecessary to accumulate wealth. If debt had been prohibited, the asset curve might have looked as indicated by the broken line. In view of time preference and, for commodities, storage costs, this would have entailed higher holding costs. Debt thus permits a lowering of asset holding costs, but it also costs interest, both interest foregone on the interest-earning assets (if any) which would otherwise have been held, and interest paid on the debt. In addition, borrowing will often involve additional transactions costs. Debt is incurred up to the point where the marginal benefit from lower asset holding ceases to outweigh the marginal interest and transactions costs.[10] This marginal consideration can be visualized in Figure 1 by moving the vertical scale upward by $1, thus cutting off a narrow strip over the length of the asset profile. This has the following consequences: 1) It saves time preference over the length of the year, reflecting the cost of accumulating a dollar of assets; 2) it saves storage costs during the time inventories are held; 3) it costs interest (paid or foregone) during the period of debt or bond holding; 4) it adds costly transactions commodities/money/bonds and money/bonds where indicated by a heavy line (note that switching from bonds to debt adds no transactions cost). This intuitive explanation of asset diversification will be made precise and more general by the formal model.

II. Individual Optimization

A. *The Constraints*

Let an individual's endowment be described by the vector $(\bar{x}_1, \ldots \bar{x}_t, \ldots \bar{x}_h)$. The task is to transform this into the consumption vector $(x_1, \ldots x_t, \ldots x_h)$ which maximizes $U = U(x_1, \ldots x_t, \ldots x_h)$, subject to linear constraints. There are three sets of constraints, relating, respectively, to commodities, bonds, and money. In each set there is one constraint for each of h days ($t = 1 \ldots h$). The following notation will be used:

---

[9] No generality is hereby lost, because there is at least one such point in the course of the year.

[10] It never pays, in a pure exchange model, to go into debt in order to finance the simultaneous holding of assets. Compared to commodities and money, bonds are a low-cost asset. It never pays to reduce a low-cost asset by holding a negative amount of it just for the privilege of holding more higher-cost assets. It also never pays to incur transactions costs for the simultaneous holding of bonds and debt.

g = goods (in physical units)
m = money (in nominal terms)
b = bonds (in pieces of bonds)
y = quantity sold
z = quantity bought
s = quantity held in stock
p = price (in terms of money)
c = transactions costs (units of goods per unit exchanged) $(0<c<1)$
γ = storage costs on goods (percent)
r = coupon on bonds (in monetary units per bond)
ρ = interest on cash balances (percent)
t = tax (lump sum, in monetary units)

The *commodity constraint* for day t states that the increase in stocks from t to t+1 is equal to the sum of the excess of the endowment over consumption and net purchases after deducting what is wasted in storage and used up in transactions of goods and bonds:

$$(1)\quad s^{g}_{t+1} - s^{g}_{t} = (\bar{x}_t - x_t) + (z^{g}_t - y^{g}_t) - \gamma s^{g}_t - c^{g}(z^{g}_t + y^{g}_t) - c^{b}(z^{b}_t + y^{b}_t)$$

Optimization will see to it that either purchases or sales (or both) are zero for both goods and bonds. With transactions costs it is wasteful to buy and sell simultaneously in the same market. The *bond constraints* say simply that the change in the portfolio equals the excess of purchases over sales:

$$(2)\qquad s^{b}_{t+1} - s^{b}_{t} = z^{b}_t - y^{b}_t$$

The *money constraints* state that cash balances are augmented by net sales of commodities and bonds, by the receipt of coupon payments on bonds, and by interest payments on cash balances, after deducting tax payments:[11]

$$(3)\quad s^{m}_{t+1} - s^{m}_{t} = p^{g}_t(y^{g}_t - z^{g}_t) + p^{b}_t(y^{b}_t - z^{b}_t) + r s^{b}_t + \rho s^{m}_t - t_t$$

To these constraints we have to add the requirement that the stocks of commodities and money and the flows of commodities and bonds be nonnegative:

$$a \geq 0 \qquad a = x,\ y^{g},\ z^{g},\ y^{b},\ z^{b},\ s^{g},\ s^{m}$$

There is no sign constraint for bond stocks, since the individual is free to go into debt.[12] We also have to formalize the stationarity requirement, according to which an individual's choice is restricted to plans resulting in the same cycle of consumption, exchange, and stocks year by year:

$$q_{nh+t} = q_t \qquad \text{for } q = x,\ y^{g},\ z^{g},\ y^{b},\ z^{b},\ s^{g},\ s^{m},\ s^{b}$$
$$n = 0 \ldots . (\mathrm{H} - 1)$$

What an individual is asked to determine is his optimal stationary plan.

This is a non-linear programming problem with a quasi-concave objective function and linear constraints. Its solution determines the consumption of goods, the purchases and sales of goods and bonds, and the stocks of goods, money, bonds, and debt for every day, given the pattern of endowments, the history of market prices for goods and bonds, time preference, transactions and storage cost rates, taxes, the coupon on bonds, and the rate of interest on cash balances.

B. *A Two-Day Illustration*

Some intuitive insight into this optimization problem (though not into portfolio diversification) can be gained from a graphical representation (Figure 2) by shortening the "year" to two "days." The individual wishes to transform his endowment, represented by P, into the consumption plan with highest utility, given trans-

---

[11] Note that taxes may vary over time; they may, for example, be zero for every day but one.

[12] A corresponding sign constraint for bonds would express the assumption that all bonds are issued by the government.

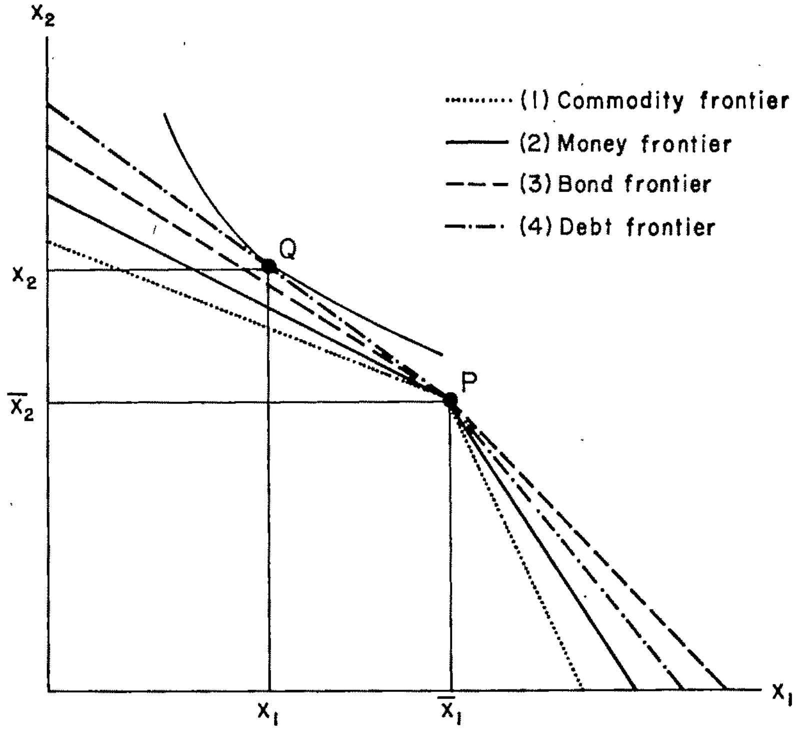


FIGURE 2

actions and storage costs and market conditions. Day 1 goods can be transformed into day 2 goods:

(a) by holding goods from day 1 to day 2;
(b) by selling goods on day 1, holding money during day 2, and buying goods on day 2;
(c) by selling goods to buy bonds on day 1, holding bonds during day 2, and selling bonds to buy goods on day 2;
(d) by borrowing to buy goods on day 2, holding a debt during day 1 of the following year, and selling goods to repay the debt on day 1 of that year.

For each sequence, the relevant constraints tell us how much x_2 can be increased by giving up a certain quantity of x_1. We thus obtain the linear segments of the commodity, money, bond, and debt frontier sloping upward from P.[18] Similar segment

---

[18] Their slopes are calculated by solving for $-dx_2/d$ from the relevant constraints, omitting those item that do not appear in the particular trading sequenc The slopes are

$$\text{(a)} \quad -\left(\frac{dx_2}{dx_1}\right)^g = 1 - \gamma$$

$$\text{(b)} \quad -\left(\frac{dx_2}{dx_1}\right)^m = (1+\rho)\,\frac{1-c^g}{1+c^g}\cdot\frac{p_1^g}{p_2^g}$$

$$\text{(c)} \quad -\left(\frac{dx_2}{dx_1}\right)^b = \frac{(1-c^g)(p_2^b+r)/p_2^g - c^b}{(1+c^g)p_1^b/p_1^g + c^b}$$

$$\text{(d)} \quad -\left(\frac{dx_2}{dx_1}\right)^d = \frac{(1-c^g)p_2^b/p_2^g - c^b}{(1+c^g)(p_1^b+r)/p_1^g + c^b}\,(1+\theta)^2$$

In sequence (d), the entries relating to day 1 of the fo lowing year have to be discounted for two days to mak them comparable with the entries for day 1 of the cu rent year in sequences (a)–(c). A comparison of the deb slope (d) with the bond slope (c) confirms nicely tha the advantage of debt, if any, is due to the time prefe ence effect, working through θ, dominating the interes effect, working through r (transactions costs being th same in both cases).

can be obtained for the reverse shift of consumption from day 2 to day 1 (of the following year), resulting in the rays sloping downward from P. Due to transactions costs, market returns, and time preference, each of the four frontiers has a convex kink at P.[14] The overall transformation frontier consists of the outwardmost segments of the commodity, money, bond, and debt frontiers. As the graph is drawn, it is efficient either to be in debt or to hold bonds during the first day. With different transactions costs, market returns, prices, and time preference, other sequences could have been efficient.[15] Once we know the transformation frontier, the optimum is determined by confronting it with the utility function. In the graph, the individual chooses Q, thereby determining all stocks and flows. If Q coincides with P, no assets are held. It is clear that in a two-day model an individual will never hold assets on both days and that his portfolio will never be diversified. With more days in the cycle, the transformation frontier will generally contain more complex sequences involving diversified portfolios.

C. *The Marginal Conditions*

More insight into the economic character of the optimal solution can be gained by deriving the Kuhn-Tucker conditions on consumption, exchange flows, and stocks.[16] They are obtained by forming the Lagrange equation

$$
\begin{aligned}
(4)\quad L = {} & U(x_1 \, . \, . \, x_h) \\
& - \sum_{t=1}^{h} \lambda_t^g \big[(s_{t+1}^g - s_t^g) - (\bar{x}_t - x_t) \\
& - (z_t^g - y_t^g) + \gamma s_t^g + c^g(z_t^g + y_t^g) \\
& + c^b(z_t^b + y_t^b)\big] \\
& - \sum_{t=1}^{h} \lambda_t^b \big[(s_{t+1}^b - s_t^b) - (z_t^b - y_t^b)\big] \\
& - \sum_{t=1}^{h} \lambda_t^m \big[(s_{t+1}^m - s_t^m) - p_t^g(y_t^g - z_t^g) \\
& - p_t^b(y_t^b - z_t^b) - r s_t^b - \rho s_t^m + t_t\big]
\end{aligned}
$$

and taking its partial derivatives. The multipliers $\lambda_t^i (i = g, m, b)$ can be interpreted as discounted marginal (flow) values of goods, bonds, and money.[17] These can be transformed into (undiscounted) marginal values $\mu_t^i = (1+\theta)^t \lambda_t^i$ to make the role of time preference explicit. The marginal values are subjective magnitudes, measured in terms of utility.

For *consumption* the marginal conditions are

$$
\begin{aligned}
(5)\quad & \left(\frac{\partial U}{\partial x_t} - \lambda_t^g\right) x_t = 0 \\
& \frac{\partial U}{\partial x_t} \leq \lambda_t^g = \left(\frac{1}{1+\theta}\right)^t \mu_t^g \\
& x_t \geq 0
\end{aligned}
$$

If, on a given day, there is any consumption, the direct marginal utility of commodities is equal to their marginal value. Whenever the direct marginal utility is lower than the marginal value, no goods are consumed. In this case, goods are valued not because of their direct utility, but because of what they bring in exchange or through storage; their value, like the value of money and bonds, is then of an indirect nature.

For *exchange flows* the economic interpretation of the marginal conditions is also straightforward. The conditions state that exchange takes place up to the point where the marginal rate of substitution between goods (or bonds) and money is equal to their price *after transactions costs*. It is because of this wedge between market prices and marginal values, due to transactions costs, that assets may be diversified. For the same reason we have to distinguish between sales and purchases. The condition for commodities, bonds, and money are given in turn.

---

[14] Convexity can be shown to follow from the conditions for a stationary-motion equilibrium. See Section IIIB below.

[15] It should be noted that there can be no equilibrium with private debt unless both borrowing and lending are efficient at the same time. Otherwise the borrowers could find no lenders or the lenders no borrowers. The bond price would shift, swiveling the two frontiers around P, until this condition is satisfied or else the supply of private debt has disappeared.

[16] Kenneth Arrow and Alain Enthoven have shown that the Kuhn-Tucker conditions are necessary and sufficient for a maximum in a problem of this type.

[17] See M. L. Balinski and Baumol, where the required qualifications are also given.

Commodity Sales

$$(6) \qquad [(1+c^g)\lambda_t^g - p_t^g\lambda_t^m]y_t^g = 0$$

$$\frac{\lambda_t^g}{\lambda_t^m} = \frac{\mu_t^g}{\mu_t^m} \geq \frac{p_t^g}{1+c^g}$$

$$y_t^g \geq 0$$

Commodity Purchases

$$(7) \qquad [(1-c^g)\lambda_t^g - p_t^g\lambda_t^m]z_t^g = 0$$

$$\frac{\lambda_t^g}{\lambda_t^m} = \frac{\mu_t^g}{\mu_t^m} \leq \frac{p_t^g}{1-c^g}$$

$$z_t^g \geq 0$$

Whenever there is positive supply or demand, the corresponding sign becomes an equality. In the absence of transactions costs we obtain the familiar condition that whenever exchange takes place the marginal rate of substitution between goods and money is made equal to the market price. With transactions costs, the marginal rate of substitution will be higher (for sales) or lower (for purchases) than the market price. If the price *after transactions costs* is lower than the marginal rate of substitution, no sales take place. If the price after transactions cost is higher than the marginal rate of substitution, no commodities are bought. There can never be supply and demand at the same time. If strict inequalities hold for both purchases and sales, no exchange will take place on that day.

The terms λ_t^m and μ_t^m represent the ***marginal flow values of money***, discounted and undiscounted. They are flow values in the sense that they express the gain in utility resulting from a free gift of a dollar, to be used up as the individual sees fit. Satiation with cash balances, in the sense in which this term is used in the optimal money supply literature, cannot mean that λ_t^m is zero. For the individual confronted with given market prices, λ_t^m would go to zero only if λ_t^g went to zero, that is, as we approach perfect bliss. As long as there are desires, a free gift of money is always welcome. For society as a whole, with market prices variable, λ_t^m goes to zero as a progressive increase in the nominal money supply, with given λ_t^g, drives prices to infinity. Neither of these two meanings of satiation is relevant for the optimality problem, which has to do with real balances in a world of continued scarcity of resources. The relevant concept will emerge presently.

Bond Sales

$$(8) \qquad (\lambda_t^b + c^b\lambda_t^g - p_t^b\lambda_t^m)y_t^b = 0$$

$$\frac{\lambda_t^b}{\lambda_t^m} = \frac{\mu_t^b}{\mu_t^m} \geq p_t^b - c^b\frac{\mu_t^g}{\mu_t^m} = p_t^b - c^b\frac{\lambda_t^g}{\lambda_t^m}$$

$$y_t^b \geq 0$$

Bond Purchases

$$(9) \qquad (\lambda_t^b - c^b\lambda_t^g - p_t^b\lambda_t^m)z_t^b = 0$$

$$\frac{\lambda_t^b}{\lambda_t^m} = \frac{\mu_t^b}{\mu_t^m} \leq p_t^b + c^b\frac{\mu_t^g}{\mu_t^m} = p_t^b + c^b\frac{\lambda_t^g}{\lambda_t^m}$$

$$z_t^b \geq 0$$

The interpretation of the bond exchange conditions is similar to that of the commodity conditions. The only difference is that transactions costs on bonds are valued at the marginal value not of bonds, but again of commodities. This is, of course, because all transactions costs are assumed to consist of commodity inputs. If, in the given period, goods are actually sold or bought, we can substitute for μ_t^g/μ_t^m from (6) or (7) to express the condition for μ_t^b/μ_t^m in terms of market data only. The higher the transactions costs, the more likely it is that strict inequalities will appear for both sales and purchases; no bond transactions can then take place. This is the "locking in" effect of transactions costs.

The most interesting marginal conditions are those on *stocks*. They relate the marginal stock value of an asset to its opportunity cost. The marginal stock value represents the gain in utility an individual derives from *holding* a unit of an asset for one day. It is measured simply by the overnight (percentage) increase in the marginal flow value of that asset. In this way, marginal stock values are precisely related to marginal flow values. It is one of the virtues of the approach presented here that the dimensional confusion about the marginal value of money and other assets can be cleared up. The opportunity cost of an asset, on the other hand, is simply its yield, reflecting storage costs, market return, and time preference, but not including capital gains or losses due to price changes. Again, the marginal conditions can be expressed in discounted or undiscounted form. It is here where the difference becomes important, because the role of time preference as a component of asset holding costs, which is implicit in the discounted formulation, becomes explicit in undiscounted terms.

Commodity Stocks

$$(10) \qquad [(1-\gamma)\lambda_t^g - \lambda_{t-1}^g]s_t^g = 0$$

$$\frac{\lambda_t^g - \lambda_{t-1}^g}{\lambda_t^g} \leq \gamma \qquad s_t^g \geq 0$$

or

$$(10') \qquad \frac{\mu_t^g - \mu_{t-1}^g}{\mu_t^g} \leq \frac{1}{1+\theta}(\theta+\gamma)$$

Condition (10′) says that the daily valuation gain on commodity stocks, expressed as a percentage, cannot exceed the opportunity cost of commodity stocks, $\theta+\gamma$, properly discounted to the same moment. Whenever the marginal stock value of inventories is higher than their holding cost, it is profitable to increase them and equilibrium cannot yet have been reached. Conversely, if the marginal stock value is lower than the holding cost, there is an incentive to reduce the stock. However, once the stock has reached zero, no further reduction is possible and the inequality persists even in equilibrium. Walras, in developing his theory of money, postulated a special *service d'approvisionnement* for inventories. The present analysis, by explicitly relating the marginal stock value of goods to their marginal flow value, makes such constructs unnecessary.

Bond Stocks

$$(11) \qquad (\lambda_t^b - \lambda_{t-1}^b + r\lambda_t^m)s_t^b = 0$$

$$\frac{\lambda_t^b - \lambda_{t-1}^b}{\lambda_t^b} = -r\frac{\lambda_t^m}{\lambda_t^b} \qquad s_t^b \gtreqless 0$$

or

$$(11') \qquad \frac{\mu_t^b - \mu_{t-1}^b}{\mu_t^b} = \frac{1}{1+\theta}\left(\theta - r\frac{\mu_t^m}{\mu_t^b}\right)$$

The marginal stock value for bonds has the same appearance as that for commodities, expressing the daily increase in the subjective value of the bond. On the side of the opportunity cost, we are not surprised to find the coupon in the place of the storage cost rate, of course with opposite sign. The coupon rate r is weighted with μ_t^m/μ_t^b, because the coupon is paid in terms of money. It should be noted that the opportunity cost of bonds is not, as in the case of commodities and money, fixed exogenously, but through μ_t^m/μ_t^b is related to the endogenous variables of the model.

The marginal condition for bonds appears as an equality, because the individual can always make the marginal stock value equal to the opportunity cost either by holding bonds or by going into debt, thus making s_t^b positive or negative. If the bond stock is negative, the right-hand side of (11′) can be interpreted as the net marginal benefit of debt; debt permits a reduction in assets and thus in time preference costs, represented by θ, but this benefit is reduced by the interest costs. If we modify the model by excluding private debt, thus leaving only government bonds, the equality is replaced by $\leq$; once the stock is reduced to zero, nothing further can be done to raise the marginal stock value to the level of the opportunity cost.

Money Stocks

$$(12) \qquad [(1+\rho)\lambda_t^m - \lambda_{t-1}^m]s_t^m = 0$$

$$\frac{\lambda_t^m - \lambda_{t-1}^m}{\lambda_t^m} \leq -\rho \qquad\qquad s_t^m \geq 0$$

$$(12') \qquad \frac{\mu_t^m - \mu_{t-1}^m}{\mu_t^m} \leq \frac{1}{1+\theta}(\theta - \rho)$$

The marginal stock value of cash balances measures the net benefit from *holding* a dollar overnight. It can be visualized as the benefit from spending the dollar today minus the sacrifice involved in not spending it yesterday, both expressed as a percentage of today's value. While the marginal flow value of money, as pointed out above, does not go to zero for a single individual short of perfect bliss, the marginal stock value may well go to zero, indicating that larger *stocks* of money promise no further benefit. This is the satiation concept that is relevant to the problem of an optimal money supply. The opportunity cost of cash balances is simply the difference between the rate of time preference and the rate of interest on money, if any. Whenever money is held, its marginal stock value is equalized to its opportunity cost. If $\rho=\theta$, satiation with cash balances is consistent with positive cash holdings. In fact, satiation balances will then be held at all times.[18]

By relating the discounted marginal stock values for goods and money, we derive

$$(13) \qquad \frac{\dfrac{\lambda_t^g - \lambda_{t-1}^g}{\lambda^g}}{\dfrac{\lambda_t^m - \lambda_{t-1}^m}{\lambda_t^m}} = \frac{-\gamma}{\rho}$$

$$s_t^g > 0, \qquad s_t^m > 0$$

This means that the ratio between the discounted marginal stock values for goods and money, *if both are held*, is equal to the ratio between the two rates of return. A similar relationship, but in absolute terms, holds for bonds and money:

$$(14) \qquad \frac{\lambda_t^b - \lambda_{t-1}^b}{\lambda_t^m - \lambda_{t-1}} = \frac{r}{\rho}$$

$$s_t^m > 0$$

These conditions are the stock counterpart to the proposition, familiar from neoclassical flow models, that the marginal rate of substitution for any two goods is equal, in equilibrium, to the ratio of their prices. They are the exact marginal conditions for portfolio diversification.

---

[18] If $\rho=\theta$, then $\mu_t^m - \mu_{t-1}^m \leq 0$ for all t. But stationarity requires that at the end of the year this marginal stock value be at the same level as at the beginning. Thus it must be equal to zero for all t.

If there is a storage cost on commodities ($\gamma > 0$), the opportunity cost of cash balances is lower than that of commodity stocks (provided $\rho \geq 0$). Nevertheless an individual may prefer to hold commodities, because a shift into money would involve transactions cost; for short holding periods this may weigh more heavily than the saving in holding costs.[19] On the other hand, the market will see to it that the opportunity cost of bonds is lower than that of cash balances for most of the time,[20] but again transactions costs may prevent a further shift into the lower-cost asset.[21]

Neoclassical models, by introducing the money stock as an argument into the utility function, usually yield a marginal condition of the type $(\partial U/\partial M)/i = \lambda$, where $\partial U/\partial M$ represents the marginal stock utility of money (in absolute terms, not in percent), λ is the marginal utility of income, and i is the rate of interest, supposed to measure the opportunity cost of cash balances.[22] This is in close analogy to (11′). We just rewrite (11′) as

$$\frac{\overset{m}{\mu}_t - \overset{m}{\mu}_{t-1}}{\theta - \rho} \leq \frac{1}{1 + \theta} \overset{m}{\mu}_t$$

The term $(\mu_t^m - \mu_{t-1}^m)$ is seen to correspond to $\partial U/\partial M$, $(\theta - \rho)$ expresses the opportunity cost of cash balances, while μ_t^m is the marginal *flow* utility of money, analogous to the marginal utility of income. Where the neoclassical models have to *postulate* a marginal stock utility of money, the present model, by a straightforward generalization, is able to *derive* it.[23]

III. Market Equilibrium

A. *The Equilibrium Conditions*

Once we know how each individual determines his optimal plan for given prices of goods and bonds, these prices can be determined by the condition that demand equal supply in all markets. The equilibrium conditions are simple. Denoting individuals by left-hand subscripts, they can be written

$$(15) \qquad \sum_i {}_i s_t^m = M, \qquad M > 0$$

$$(16) \qquad \sum_i {}_i s_t^b = B, \qquad B \geq 0$$

$$(17) \qquad T = \rho M + rB$$

$$(18) \qquad \sum_i {}_i t_t = T$$

Equation (15) says that on each day the total demand for money must be equal to the given money stock M; equation (16) requires the same for bonds; equation (17) is the government budget constraint, stating essentially that the government's only activity consists in levying taxes to pay interest on cash balances and consols. M, B, ρ, and r are assumed to be constant over time and thus carry no time subscript. Equation (18) finally states that individual tax burdens must add up to the total required by given M, B, ρ, and r; in principle ${}_i t_t$ can be freely chosen for all but one individual. There are no equilibrium conditions for the demand and supply of commodities; if there is equilibrium for bonds and money, there is equilibrium for commodities by virtue of Walras' Law. It can easily be ascertained that an equiproportionate increase in M, r, and all ${}_ih_t$ is neutral in the sense that it results in an equiproportionate change in all commodity prices, bond prices, and money stocks, while the real variables remain unchanged.[24] The quantity theory holds.[25]

---

[19] For just two days, the relevant comparison is between slopes $-(dx_2/dx_1)^g$ and $-(dx_2/dx_1)^m$ in fn. 13.

[20] If it did not, nobody would be willing to hold bonds, and there would be a point at which it would become profitable to borrow money just to hold it idle.

[21] The exact condition for this to happen in the two-day case is obtained by comparing slopes $-(dx_2/dx_1)^m$ and $-(dx_2/dx_1)^b$ in fn. 13.

[22] See, for example, Paul Samuelson, p. 9.

[23] A similar expression for the marginal utility of money is derived by Grandmont and Younes, p. 154, by replacing the postulate of a direct utility of money by the postulate of a Cambridge-type demand function for money.

There is, of course, no general rule that money must always have a positive value in the economy of the present model. It is logically possible to conceive of a case in which each individual's endowments and tastes are synchronized in such a way that no assets and thus no cash balances are ever held. If assets are held, individuals may still be identical; as a consequence, there would be no incentive for exchange and thus no demand for a medium of exchange. Finally, even if individuals are sufficiently different, transactions costs on bonds and storage costs on goods may be so low and transactions costs on goods may be so high, that no assets are ever held in the form of money. Without transactions costs on bonds, bonds would circulate in the place of money. On the other hand, by selecting appropriate values for transactions and storage costs we can always produce asset diversification.[26] Whether a monetary economy is viable depends not on one simple condition, but on the interplay of all the factors determining equilibrium.

B. *Limiting Conditions on Market Returns*

Equilibrium requires that individuals be satisfied with finite holdings of assets and debt. We can certainly visualize market returns so high that further accumulation always pays; asset demand then becomes infinite. Equilibrium thus imposes limits on the admissible rates of asset returns.

For commodities and money, these limits are easily derived from the stationarity conditions. Stationarity is inconsistent with negative values of $(\mu_t^g-\mu_{t-1}^g)$ and $(\mu_t^m-\mu_{t-1}^m)$ for all t, because the marginal flow values would then decline from year to year. But according to the stock conditions for commodities (10′) and money (12′) these expressions would indeed be negative for all t if the respective opportunity cost, appearing on the right-hand side of (10′) and (12′), were negative. It follows that the opportunity costs must be nonnegative. There can be no positive net yield on inventories and cash balances.

The economic interpretation of this argument is straightforward. Consider money. A nonnegative opportunity cost means $\rho/\theta \leq 1$. The left-hand side measures the present value of a dollar, held forever. The right-hand side is simply the money price of a dollar. As long as the condition is satisfied, it does not pay to hold dollars forever; it only pays to hold them temporarily, when the market return is augmented by exchange services. However, if the market return on money exceeds the rate of time preference, it becomes profitable to hold cash balances permanently without ever spending them, and if this pays for one dollar, it pays for any amount. There can be no equilibrium. At the exact point where $\rho=\theta$, the economy just satiates itself with cash balances, the marginal stock values of money go to zero, the marginal flow values of money are equalized over time, and the stock demand for cash balances becomes indeterminate. The same considerations can be applied to commodity stocks. With storage costs, the limiting conditions are, of course, always satisfied. If storage costs are replaced by positive physical growth rates, resulting in a positive market return, then the limiting condition appears as the familiar rule that the growth rate must not exceed the rate of time preference.

---

[24] It can be checked that the constraints (1)–(3) and the marginal conditions (5)–(12) are not affected by such a change. The marginal flow value of money, λ_t^m, moves in inverse proportion, of course.

[25] This neutrality property depends crucially on the assumption that all money is token money.

[26] Numerical examples for a basically similar, but less general model were given in the author's 1970 paper.

The opportunity cost for bonds (or debt) differs from that for commodities and money to the extent that, in view of μ_t^m/μ_t^b, it can fluctuate over time. This means that negative opportunity costs for certain days, associated with falling μ_t^b, may be compatible with stationarity, provided they are accompanied by positive opportunity costs for other days, associated with rising μ_t^b. However, there must still be limits which bond prices cannot exceed without making equilibrium impossible. A lower limit is derived by considering bonds as an asset. To exclude infinite accumulation, the price of a bond, including transactions costs, must be at least as high as its present value, if held forever. This results in the condition

$$p_t^b + c^b \frac{\mu_t^g}{\mu_t^m} \geq \frac{r}{\theta} \tag{19}$$

The upper limit is relevant for bonds considered as a debt. To exclude an infinite amount of debt, the proceeds from selling a bond, net of transactions costs, must be no higher than the present value of the debt, maintained permanently, or

$$p_t^b - c^b \frac{\mu_t^g}{\mu_t^m} \leq \frac{r}{\theta} \tag{20}$$

These conditions must be satisfied for each individual.

While the limiting condition for money constrains the government in fixing the rate of interest on cash balances, the corresponding conditions for bonds are not a constraint on the coupon paid by the government or private borrowers. The amount of the coupon is economically irrelevant. What is limited are the possible highs and lows of the bond price. The market will automatically see to it that the limits are not exceeded. Suppose the bond price is set arbitrarily below the lower limit. Bond demand will become infinite, while the supply is finite at the level fixed exogenously by the government. It follows that the assumed price cannot be an equilibrium price; it will be bid up until it satisfies the limiting condition. The reverse is true if the bond price is set above its upper limit: there will be an infinite supply of bonds, but no demand, and the excess supply will drive the price down again.

IV. The Rate of Interest on Money and the Optimal Money Supply

Milton Friedman has proposed that the government provide for a return on cash balances equal to the rate of time preference. This proposition may fairly be called the Chicago Rule.[27] So far, its general equilibrium analysis has been fragmentary.[28] The model presented in the preceding sections can be used to shed light on its implications. It will be assumed that the return on cash balances is provided, not by deflation, but by interest payments on cash balances, financed by lump sum taxes. It will also be assumed that θ is equal for all individuals.

A. *The Demand for Cash Balances*

The first question is whether the demand for real cash balances as a function of their opportunity cost $(\theta - \rho)$ can be shown to be negatively sloped. For a year of many days it is difficult to give this question a precise meaning; real cash balances generally change from day to day, and an increase in ρ may well result in an increase in s_t^m for some days, but a decrease for others. I thus concentrate on the two-day case, considering a single individual. Any increase in real balances, achieved by the individual through an increase in nominal balances at given prices, will be achieved by the economy as a whole by a fall in prices at a given nominal money supply. It will be shown that the demand for real balances indeed rises as $(\theta-\rho)$ falls, *provided the rise in ρ is compensated by taxes.*

---

[27] See also Harry Johnson.

[28] See Hahn's critique (1971a). Grandmont and Younes take up the problem, but they make the demand for money simply a function of purchases and sales. They also do not consider credit.

We first assume that there are no bonds and that transactions and storage costs make it inefficient ever to hold commodity stocks. The transformation frontier of Figure 2 thus coincides with the money frontier; money is the only efficient asset. Figure 3 represents this case. The inner frontier with optimum Q depicts the initial situation. The demand for cash balances can be measured by $\bar{x}_1 - x_1$.[29] An uncompensated increase in ρ expresses itself by an outward rotation of the transformation frontier around P. If the new optimum is at a point like Q', cash balances rise. However, if the new equilibrium is at a point like Q'', the demand for cash balances falls.[30] The reason is that an income effect of uncertain sign is added to the positive substitution effect. If neither x_1 nor x_2 is inferior, the income effect will in fact be negative, and it may easily dominate the substitution effect. This ambiguity is familiar from the effect of changes in the rate of interest on saving.

The introduction of a compensating lump sum tax removes this ambiguity. The tax $t_2^* = \rho s_2^m$, necessary to cover the interest payments,[31] is reflected in Figure 3 in a downward shift of the transformation frontier by $(1-c^g)\ t_2^*/p_2^g$, which is the amount of day 2 consumption which could have been bought with the tax payment. The resulting equilibrium Q''' is certain to be to the left of Q.

If assets are initially held in the form of bonds and/or commodity stocks, the interest elasticity of the demand for money is further increased. In terms of Figure 2, we now start with a money frontier still inside the bond and/or commodity frontier. As ρ increases, the faces of the money frontier will swivel outward, until a point is reached at which it becomes efficient to switch from bonds or commodities into cash. The own-interest elasticity of the demand for cash balances thus depends both on the reaction of total assets and the substitution of money for other assets.

At the point where $\theta - \rho = 0$, every individual is satiated with cash balances. According to (12′), $\mu_t^m - \mu_{t-1}^m = 0$ for all t; the marginal stock values of money go to zero, while the marginal flow values are equalized over time. At the same time, the demand for real balances becomes indeterminate in the sense that individuals are indifferent to holding even higher money stocks, no matter what their amount. For the economy as a whole this means that the price level is indeterminate. The economy is in a liquidity trap, a general equilibrium counterpart to the Keynesian liquidity trap. This indifferent equilibrium is precarious, however. Any further increase in ρ would violate the stationarity condition given in Section IIIB. A stationary-motion equilibrium would no longer be possible. The stock demand for money

---

[29] From constraints (1) and (3) we derive

$$s_2^m = \frac{p_1^g}{1+c^g}(\bar{x}_1 - x_1) = \frac{1}{1+\rho}\,\frac{p_2^g}{1-c^g}(x_2 - \bar{x}_2)$$

[30] In view of the simple linear relationship

$$\frac{\lambda_2^m - \lambda_1^m}{\lambda_2^m} = 1 - \left(\frac{p_2^g}{p_1^g}\cdot\frac{1+c^g}{1-c^g}\right)\frac{\lambda_1^g}{\lambda_2^m}$$

the steeper slope of the indifference curves at the new optimum is, for given prices and transactions costs, an exact measure of the decline in the (discounted) marginal stock value of money.

[31] For the economy as a whole, the nominal money stock does not change, increases in real balances being achieved by price declines. For a representative individual, the tax thus has to cover the additional interest on the cash balance held *before* the increase in ρ.

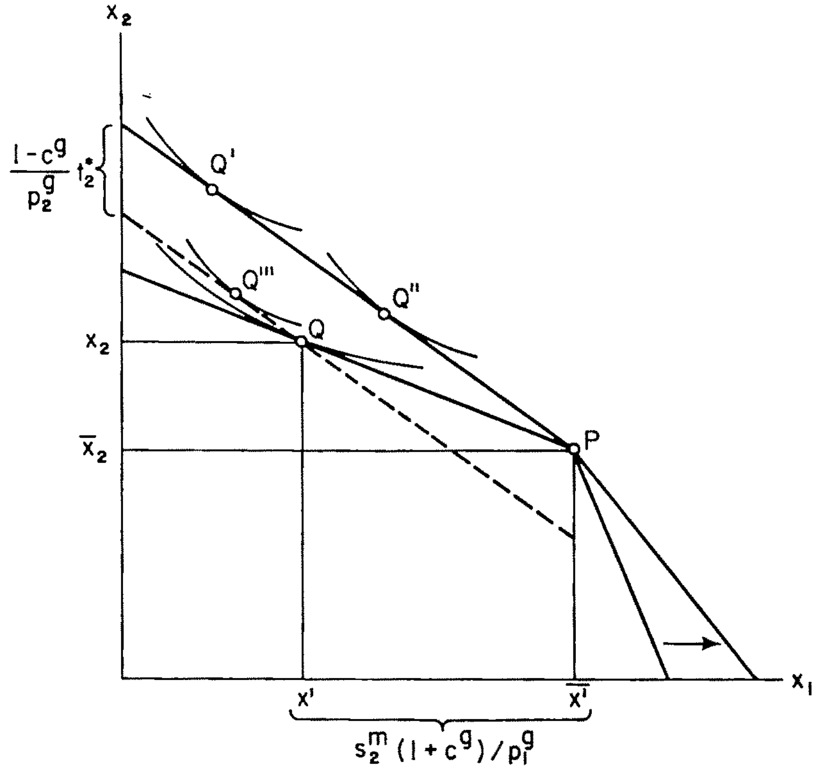


FIGURE 3

would become insatiable, and no price level above zero could provide equilibrium.[32] The Chicago Rule seems to place the economy on a knife-edge between scarcity of money and a pathological addition to higher and higher cash balances.

B. *Welfare*

The second question is whether the Chicago Rule indeed offers an increase in welfare. It offers an unambiguous increase in welfare if, and only if, an equilibrium with $\rho=\theta$ is Pareto superior to an equilibrium with $\rho<\theta$, in the sense that an increase in ρ from $\rho<\theta$ to $\rho=\theta$, all other exogenous variables remaining the same, makes some people better off without making anybody worse off.[33] Can this be shown to be the case?

Consider first a single individual confronted with given patterns of commodity and bond prices. The equalization of the marginal flow values of money means that an individual's utility can no longer be increased by offering him early a dollar he would otherwise have received later. Everybody will now behave *as if* the money constraints (3), one for each day, had been replaced by an annual budget constraint. The timing of cash flows no longer matters. This is a clear benefit, illustrated in Figure 3 by the fact that Q' and Q'' are on a higher indifference curve than Q.[34] The costs consist, of course, of the tax t_2^*. A compensated increase in ρ is analogous to the elimination of an excise tax on cash balances, financed by lump sum taxes. By the well-known argument, this makes the consumer better off; Q''' is necessarily on a higher indifference curve than Q. If prices remained unchanged, cash satiation would indeed make the economy Pareto superior.

---

[32] This is similar to the indeterminacy and nonexistence of equilibrium found by Grandmont and Younes. Feige and Parkin, p. 348, have no counterpart of this result, because they assume total wealth to be given exogenously.

[33] This question should be distinguished from the question of Pareto optimality. As is well known, Pareto optimality of state A is neither necessary nor sufficient for its Pareto superiority over some other state B. Suppose it can be shown that Pareto optimality requires $\rho=\theta$. It does not follow that, starting from a given state B with $\rho<\theta$, an increase in ρ up to $\rho=\theta$ results in a state which is Pareto superior to B. In the present context it is Pareto superiority that counts.

However, for the economy as a whole, prices cannot remain unchanged. In fact, if some prices did not decline, there could be no increase in real balances. What is more important, relative prices will also change. The conditions for a neutral change in the price level were given in Section IIIA. A compensated increase in ρ does not satisfy these conditions. It will thus affect relative commodity prices for different days of the year as well as the interest rates on bonds.[35] Nothing general can be said about the way relative prices will change. Most likely, some individuals will make additional gains, while others will lose. It would be an unlikely coincidence if some did not lose more from these price changes than they would have gained at unchanged prices. By the criterion of Pareto superiority, therefore, the Chicago Rule cannot be shown to produce an improvement in welfare.[36]

In principle, it would be possible t modify the array of lump sum taxe (possibly changing some into subsidies) i such a way that, without a further chang in aggregate T, all losers from the Chicag Rule are at least compensated by th winners. However, as long as we do no have a workable compensation schedul and have, in fact, no idea where we shoul look for the gains and losses, this is doubtful consolation.

C. *Bonds and Private Debt*

The final question concerns the implications of the Chicago Rule for the deman for bonds and the supply of private debt With cash balances having the highes yield compatible with equilibrium, i seems intuitively plausible that there ca be no incentive ever to exchange money fo bonds. One would also suspect that ther is no motive ever to go into debt. After al why should anybody borrow if he i satiated with cash at all times? These conjectures can indeed be shown to be correct

Consider first the two-day case. B comparing the slopes of the frontiers fo money and bonds, given in Section IIB for $\theta = \rho$ and substituting for the bon price (after transactions costs) from (19 and (20), it can be shown that the bon frontier cannot lie outside the mone frontier. To hold cash is always at least a good as to hold bonds. By making a simila comparison for money and debt, it ca also be shown that the debt frontier canno lie outside the money frontier. In order t shift consumption over time, it is at leas as good to accumulate the excess supply i cash as to cover the excess demand b borrowing. In the case most favorable t bonds (or debt), an individual would b just indifferent between holding cash and holding bonds (or going into debt).

---

[34] This does not mean, of course, that the shifting of resources over time is no longer costly. If the store of value consists of commodities, there are storage costs; if it consists of money, there are transactions costs on the exchange of commodities for money. The transactions frontier in Figure 2 does not become a straight line.

[35] In a multicommodity model, there would also be changes in relative prices for different commodities.

[36] In their statements about social welfare and Pareto optimality, Feige and Parkin go beyond what their model is able to support. The main reasons are that 1) they take no account of the utility of those economic agents (presumably firms) who hold those cash balance and bonds the representative household does not wan to hold at any given time, and 2) prices and interes rates are assumed to be arbitrarily fixed.

This argument can easily be generalized to n days. According to (19), the buyer of a bond pays (after transactions costs) *at least* r/θ. After n days, he can sell the bond, according to (20), for *at most* r/θ (after transactions costs). While he may make a capital loss, he cannot make a capital gain. In the best possible case, the stream of his costs and revenues will thus be $-r/\theta$, r, $r, \ldots, r+r/\theta$. If he "invests" his funds in cash, the corresponding stream is $-r/\theta$, $(r/\theta)\rho, \ldots, (r/\theta)\rho+r/\theta$. In view of $\theta=\rho$, this is identical with the best possible bond stream. In an economy satiated with cash balances, it never pays to trade bonds. If bonds are exogenously supplied by the government, they must, of course, be held. They will be held indifferently by those who happen to have them, the market pushing the rate of interest on bonds to the level of ρ. The distribution of the bonds is as indeterminate as the amount of real balances, and they are perfect substitutes for money as stores of value. For all economic purposes, they could as well be abolished.[37]

For debt the generalization is equally straightforward. Consider a marginal reduction in the amount of debt, accompanied by an equal increase in the level of cash holdings during the rest of the year.[38] This saves transactions costs on the selling and buying of bonds. At the same time, there is an increase in time preference costs over the whole year. However, this is just matched by the higher interest income over the year, consisting partly of interest earned on the additional cash balances, partly of interest saved on the debt. As long as there is any debt, a marginal shift from debt to cash holding is always advantageous. Under the Chicago Rule there is no room for private debt. Figuratively speaking, the "*Lebensraum*" of the credit market is the margin between the rate of time preference and the rate of interest on money. If this margin is compressed to zero, private credit will be squeezed out of existence.

V. Concluding Remark

By introducing transactions costs into a general equilibrium model of stationary motion, it has been possible, I believe for the first time, to explain the stock demand for inventories, cash balances, government bonds, and private debt without assigning them direct utility and without any particular assumptions about terminal stocks. The analysis was subject to two principal limitations. The first is the neglect of credit transformation and financial intermediation. To correct it, we would have to introduce various types of interest-bearing assets, each with different characteristics. Financial intermediaries would then engage in transforming one type of asset into others, of course at the cost of resource inputs. This would open the way for a general equilibrium analysis of banking. The second limitation is the lack of a detailed analysis of transactions costs. In particular, there is no analysis of the relationship between the resources used up in transactions and uncertainty. This does not reflect a lack of appreciation for the economic significance of search and information costs. It rather reflects the tactical judgement that, at the present stage of our knowledge, search processes are difficult to incorporate into a general equilibrium analysis and that the chance for progress is better if we concentrate on a world with perfect information. There is, of course, also the more fundamental difficulty of the lack of a satisfactory theory of time preference. This is a difficulty, however, for which a solution cannot fairly be expected from monetary theory.

---

[37] In the model of Feige and Parkin, money and bonds can coexist even with an "optimal" interest rate of cash balances, if the holding costs on money exceed those on bonds (equation (35)). In the present model, these holding costs are equal, both being represented by time preference. There is thus no conflict between the respective results. Of course, for models which are so different such comparisons may be questionable to begin with.

[38] In Figure 1 this corresponds to a marginal downward shift in the vertical scale.

REFERENCES

K. J. Arrow and A. C. Enthoven, "Quasi-Concave Programming," *Econometrica*, Oct. 1961, *29*, 779–800.

M. L. Balinski and W. J. Baumol, "The Dual in Nonlinear Programming and Its Economic Interpretation," *Rev. Econ. Stud.*, July 1968, *35*, 237–56.

W. J. Baumol, "The Transactions Demand for Cash: An Inventory Theoretic Approach," *Quart. J. Econ.*, Nov. 1952, *66*, 545–56.

K. Brunner and A. H. Meltzer, "The Uses of Money: Money in the Theory of an Exchange Economy," *Amer. Econ. Rev.*, Dec. 1971, *61*, 335–49.

E. L. Feige and M. Parkin, "The Optimal Quantity of Money, Bonds, Commodity Inventories, and Capital," *Amer. Econ. Rev.*, June 1971, *61*, 335–49.

M. Friedman, *The Optimum Quantity of Money and Other Essays*, Chicago 1969.

J. Grandmont and Y. Younes, "On the Efficiency of a Monetary Equilibrium," *Rev. Econ. Stud.*, Apr. 1973, *40*, 149–65.

F. H. Hahn, (1971a) "Professor Friedman's Views on Money," *Economica*, Feb. 1971, *38*, 61–80.

———, (1971b) "Equilibrium with Transactions Costs," *Econometrica*, May 1971, *39*, 417–39.

———, "On Transactions Costs, Inessential Sequence Economies and Money," *Rev. Econ. Stud.*, Oct. 1973, *40*, 449–61.

W. P. Heller, "Transactions with Set-Up Costs," *J. Econ. Theory*, June 1972, *4*, 465–78.

J. R. Hicks, "A Suggestion for Simplifying the Theory of Money," *Economica*, Feb. 1935, *2*, 1–19.

H. G. Johnson, "Is There an Optimal Money Supply?," *J. Finance*, May 1970, *25*, 435–42.

E. Karni, "Transactions Costs and the Demand for Media of Exchange," *Western Econ. J.*, Dec. 1972, *10*, 71–80.

M. Kurz, "Equilibrium with Transactions Costs and Money in a Single-Market Exchange Economy," *Econ. Theory*, Apr. 1974, *7*, 418–52.

J. Marschak, "The Rationale of the Demand for Money and of 'Money Illusion,'" *Metroeconomica*, Aug. 1950, *2*, 71–100.

J. Niehans, "Money in a Static Theory of Optimal Payments Arrangements," *J. Money, Credit, Banking*, Nov. 1969, *1*, 706–26.

———, "Die Geldnachfrage in einer dynamischen Optimierungstheorie des Zahlungssystems," *Schweiz. Z. Volkswirtsch. Statist.*, June 1970, *106*, 129–48.

———, "Money and Barter in General Equilibrium with Transactions Costs," *Amer. Econ. Rev.*, Dec. 1971, *61*, 773–83.

———, "Veendorp on Optimal Payments Arrangements," *J. Money, Credit, Banking*, Feb. 1973, *5*, 213–14.

J. M. Ostroy, "The Informational Efficiency of Monetary Exchange," *Amer. Econ. Rev.*, Sept. 1973, *63*, 597–610.

P. A. Samuelson, "What Classical and Neoclassical Monetary Theory Really Was," *Can. J. Econ.*, Feb. 1968, *1*, 1–15.

K. Sontheimer, "On the Determination of Money Prices," *J. Money, Credit, Banking*, Aug. 1972, *4*, 489–508.

R. M. Starr, "Equilibrium and Demand for Media of Exchange in a Pure Exchange Economy with Transactions Cost," Cowles Foundation disc. pap. no. 310, Yale Univ., June 1971.

D. Starrett, "Inefficiency and the Demand for 'Money' in a Sequence Economy," *Rev. Econ. Stud.*, Oct. 1973, *40*, 437–48.

J. Tobin, "The Interest-Elasticity of Transactions Demand for Cash," *Rev. Econ. Statist.*, Aug. 1956, *38*, 241–47.

E. C. H. Veendorp, "Money in a Static Theory of Optimal Payments Arrangements —A Comment," *J. Money, Credit, Banking*, Aug. 1972, *4*, 437–40.

L. Walras, *Éléments d'économie politique pure*, 4th ed., Paris 1900.

Socially Optimal Product Differentiation

By KELVIN LANCASTER*

This paper sets out to investigate the extent to which we can reach broad general conclusions concerning the social optimality of different degrees of product differentiation. The paper is concerned with consumer preferences and production conditions only, and it abstracts from such problems as search and information costs, and disutilities of uncertainty or consumer confusion in the face of variety. It examines the problem of optimal variety in a world in which every consumer knows exactly what he prefers and exactly how to achieve personal optimality in the face of the constraints upon his actions.

Product differentiation exists when, within a group of goods so similarly related to consumers that they can be considered to form a product class, there is a variety of similar but not identical goods. The theory of product differentiation has been historically associated with the theory of monopolistic competition,[1] and has been analyzed primarily from the point of view of the firm. Although some social policy conclusions have been drawn from this approach to the subject, there does not exist a firmly based analysis leading to an answer to the question, "how many different product variants should society provide?"

One cannot go very far in answering this question within the framework of conventional or "direct" consumer theory, in which preferences are assumed to be given directly in terms of goods. For this reason, I shall turn to the "characteristics" or "indirect" analysis of consumer behavior, based on my earlier work (1966, 1971), in which the consumer is assumed to derive his actual utility or satisfaction from characteristics which cannot in general be purchased directly, but are incorporated in goods. The consumer obtains his optimum bundle of characteristics by purchasing a collection of goods so chosen as to possess *in toto* the desired characteristics.

Use of the characteristics framework provides a clear definition of a product class (those goods possessing a particular set of characteristics) and permits quantitative definition and measurement of product differentiation (by comparing the proportions in which the various characteristics are possessed by different goods within the product class), both of which properties are necessary for the analysis of optimum product variety. The characteristics framework enables us to cast the problem into a spatial setting and proceed along lines reminiscent of the pioneer work by Harold Hotelling.

The structure of the paper is as follows. First, in Section I, I shall set up some special tools of analysis that will be basic to the remainder of the analysis. Then, in Section II, I shall state and prove two basic theorems concerning optimal product differentiation and returns to scale, after which Section III examines the conditions that must be satisfied by an optimal choice for the number and type of prod-

---

* Professor of economics, Columbia University. Work for this paper was supported by the National Science Foundation (DA-39496). I wish to thank the referee for his extremely valuable comments.

[1] Chamberlin's lifelong and often pathetic struggle to retain sole title to imperfect competition should not blind us to the fact that, however the honors are to be shared between Joan Robinson and himself on the general theory of the subject, there is no doubt that it was Chamberlin who first raised the problem of the degree of product differentiation.

ucts. In Section IV, I shall state and prove four theorems concerned with optimal pricing and the effects of imperfect competition and, finally, Section V summarizes the results with special reference to policy conclusions.

The analysis will be carried out for product classes defined on two characteristics only. It will be apparent from the analysis itself and the results obtained that generalization to any number of characteristics is simply a question of additional arithmetical complexity, a complexity that does not seem justified in a pioneer investigation of the problem.

I. Tools of Analysis

A. *The Production-Consumption Link*

Consumers derive their ultimate utility or welfare from characteristics which in turn are obtained from the specific product differentiates which are available. Each product is assumed to possess those characteristics in fixed proportions. The product differentiates are themselves no more than a *transfer mechanism* by which fixed bundles of characteristics are assembled at the production end and then made available to ultimate consumers, the goods playing the role of intermediaries rather than being either primary resources or ultimate objects of consumption. If we view the system as a whole, consumer welfare is determined by the characteristics available for consumption, while the ultimate constraints are those on resources, the two linked by the transfer through goods. The transfer mechanism depends on both the way in which the resources may be used to produce goods having characteristics in different proportions and on the way in which the specific bundles of characteristics so produced are related to consumers' preferences as between all possible characteristics bundles.

For a given level of resources, the level of welfare that can be attained by the various consumers will depend on:

1) The production conditions that determine how much of each characteristic can be supplied from given resources, when embodied in a good with specific characteristics proportions.

2) The preferences of the consumers, which determine the relative welfare levels associated with various bundles of characteristics.

3) The consumption process, which determines what characteristics combinations the consumer can actually obtain from different collections of goods.

4) The number and types of goods that determine the transfer link between production and consumption.

In this paper we shall take the production and consumption possibilities as given, and be concerned primarily with the transfer mechanism. It is obvious that transfer can be efficient or not, depending on the choice of the good used for the transfer. Suppose, for example, that existing resources can be used to produce a unit of either good G_1 (embodying 2 units of characteristic z_1 and 1 of z_2) or good G_2 (1 unit of z_1, 2 of z_2). If there is a single consumer whose preferences are for high z_1-content, a lower welfare level will be attained by producing G_2 than by producing G_1. In this case, G_2 represents inefficient transfer relative to G_1, but if the consumer's preferences are biased toward z_2, the relative efficiencies of the two transfer modes will be reversed. And what if there are some consumers with preferences biased towards z_1, some towards z_2, and others in between? That is the essence of the problem we shall be solving.

B. *Production Conditions*

1. *Differentiation Possibilities*

We assume that it is possible, in principle, to produce goods having all possible ratios of the two characteristics, so that the producer can plan to produce a good anywhere in the characteristics spectrum. Having chosen the proportions in which his particular product will contain the two characteristics, there will be a unique maximum quantity of those characteristics that can be produced with given resources when incorporated in that particular good. For a given level of resources, we can take any ratio of the two characteristics, determine the maximum output of the good with that specification from the given resources and thus the maximum quantities of the two characteristics that can be produced from the resources when embodied in the appropriate good. The set of all characteristics combinations producible from a given level of resources by incorporation in a good can be plotted as a curve in characteristics space. We shall refer to this curve as a ***product differentiation curve*** and abbreviate it to *PDC*. Of particular interest will be the *PDC* corresponding to unit resource use, the ***unit PDC***.

Assume the *PDC* has geometric properties similar to those of the conventional transformation curve[2] or production possibility curve, namely: 1) it is continuous; 2) it slopes downward from left to right; 3) it is either a straight line or a curve which is concave toward the origin.

I shall make the following additional assumption: 4) *PDC*s for different resource levels are positive homothetic expansions or contractions of the ***unit PDC***. Figure 1 illustrates the assumed properties of the *PDC*s.

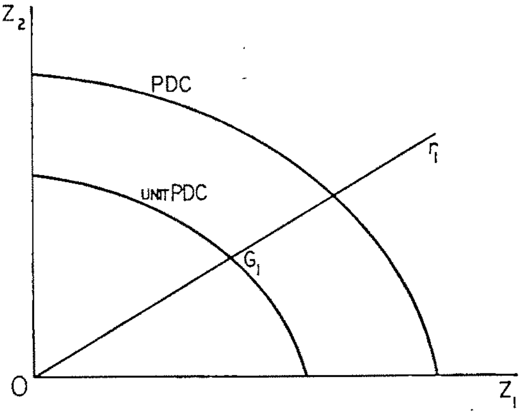


FIGURE 1

Note that I assume only a homothetic relationship between *PDC*s for different resource levels and particularly avoid any suggestion of linear homogeneity, which would imply constant returns to scale. Most of this paper is devoted to production under ***nonconstant*** returns to scale, but restricted to the homothetic case which implies the same returns to scale properties for all goods in the product class.[3]

2. *Measurement and Comparability*

The greatest single obstacle in the path of formal analysis of product differentiation is that of making quantitative comparisons between goods which are not identical. Since prices are endogenous, monetary measures cannot be used. One can only choose between comparing goods in terms of their final utility or their resource content. Since it is essential to the

---

[2] Although I referred to the *PDC* as a transformation curve in earlier versions of this paper, this is an inappropriate term. We would have a true transformation curve in characteristics space if the two characteristics were produced independently of each other, so that the curve would represent all combinations of the two separate characteristics that could be produced with given resources, each characteristic produced by its own production function with part of the total resources. Here we are concerned with the possible combinations of the two characteristics produced by incorporation in a single good. To use an analogy from high school chemistry, the regular transformation curve is concerned with ***mixtures*** of characteristics, the *PDC* with ***compounds*** made up of those characteristics.

[3] This is clearly a restrictive assumption, since it is reasonable enough to suppose that there are cases in which the returns to scale properties differ between goods in the same product class. Many of the observed differences in productivity between goods in the same product class, such as between mass-produced and custom-built automobiles, are differences in *scale* but not necessarily in the degree of *returns to scale*.

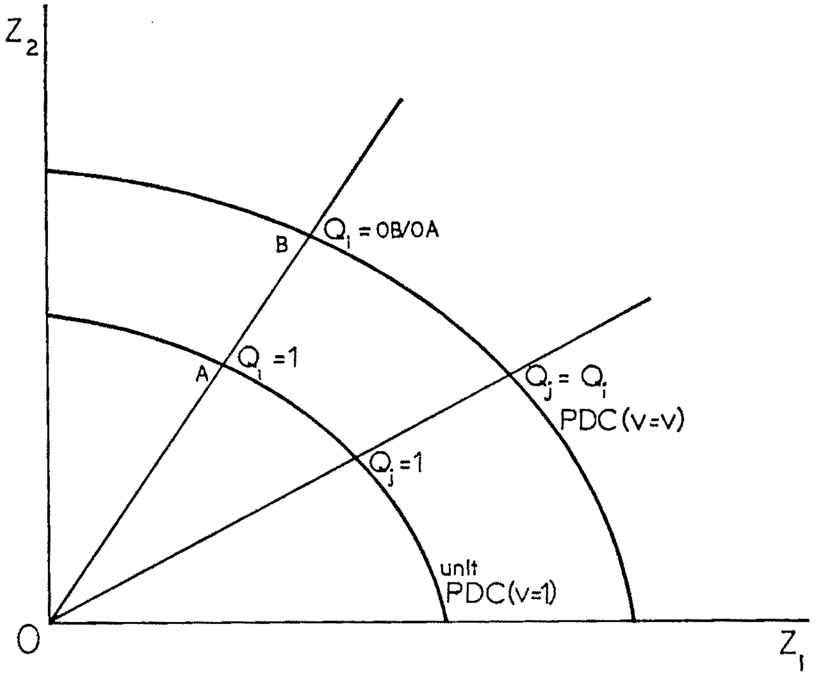


FIGURE 2

later analysis that there be many consumers with differing preferences, final utility measures are quite unsuitable. I shall, therefore, use a modified input measure that requires only that all firms face the same *PDC*s, an assumption that would be made in any case.

I shall therefore define and measure goods in the following way:

1) A good is defined by its characteristics ratio, goods with identical characteristics ratios being identical goods.

2) We bring *different* goods to the same measure by defining the unit quantity of any good to be that quantity which can be produced with unit resources.

3) Different quantities of the same good are scaled in proportion to the content of either characteristic (since the characteristics are in fixed proportions) relative to the content of that characteristic in a unit of the good as defined above.

4) As a consequence of 1)–3), plus the assumed homotheticity of the *PDC*s, quantities of different goods will receive the same measure if and only if those quantities require the same level of resources. Figure 2 illustrates the measurement system.

Note that a unit of a good is defined as requiring a unit of resources, but q units of a good do not necessarily require q units of resources since there may not be constant returns to scale. Resource content is used to relate quantities of *different* goods but quantities of the same good are scaled linearly from the resource content. Given the *unit PDC*, the characteristics content of unit quantities of different goods can be read straight off the curve and other quantities by linear scaling from these. Thus quantities of different goods are directly comparable, are in equivalent units, and can be arithmetically combined.

3. *The Input Function*

Once the choice has been made to produce a particular good with a specific characteristics ratio, there will be a defined functional relationship between the quantity of that good and the inputs required. It is convenient to use this production function in the inverse form

$$v_i = F(Q_i)$$

where v_i is the resource requirement[4] for the quantity Q_i of the good whose identity is defined by characteristics ratio $r_i (= z_2^i/z_1^i)$. I shall refer to $F(Q_i)$ as the ***input function***, to stress its inverse nature.

Due to the way in which we have defined our quantity measure and to the assumed homotheticity of the *PDC*s, F has the following very important property: ***The input function is the same for all goods. That is, the functional relationship between*** v_i ***and*** Q_i ***will be the same as the functional relationship between*** v_j ***and*** Q_j, ***for all*** i, j. (We assume, of course, that i, j are in the same product class.)

Since the input function is the inverse of a production function, $F(Q)$ increases in proportion to Q for constant returns to scale, less than in proportion for increasing returns to scale, and more than in proportion for decreasing returns. Since we shall be interested in variations in the degree of returns to scale, we note that $F(Q')/F(Q) > 1$ for all $Q' > Q$ but that the ratio approaches unity as the degree of increasing returns to scale increases without limit and the ratio increases without limit as the degree of returns to scale becomes more and more decreasing.

C. *Consumption*

1. *Consumers*

Consumers have preferences defined on characteristics. Preferences with respect to goods are indirect and derived from preferences on characteristics. These are assumed to have the same general properties with respect to characteristics as conventional preferences with respect to goods. They may be represented by preference maps with indifference curves of the conventional kind, except that we do not rule out indifference curves which are linear or piecewise linear. We shall be more interested than usual in the elasticity of substitution (curvature of the indifference curves) and sometimes concern ourselves with indifference curves showing infinite elasticity of substitution (straight lines) or zero elasticity (fixed proportions). In the latter case excess amounts of either characteristic beyond the appropriate fixed proportion will have no effect, so that zero elasticity indifference curves can be considered to extend vertically and horizontally away from the point of optimum proportions.[5]

We shall generally assume that the population consists of a very large number of consumers with different preference patterns, so that there is a continuous spectrum of preferences. Later, we shall refer to a "uniform" distribution of preferences, a term that will be defined in the appropriate context. We shall also make a special assumption about the distribution of preferences, which we introduce at the appropriate point.

2. *The Consumption Process*

Consumption involves the extraction of the characteristics embodied in the goods at the production end. There are two major technical possibilities: we may have *combinable* consumption in which goods may be combined in the consumption

---

[4] We can take the "resource" to be a single input of the dollar value of a resource mix, resource prices being taken as exogenous to the sector under consideration.

[5] Implying free disposal of the surplus characteristic, or "open satiation" as discussed in the author (1971, ch. 9).

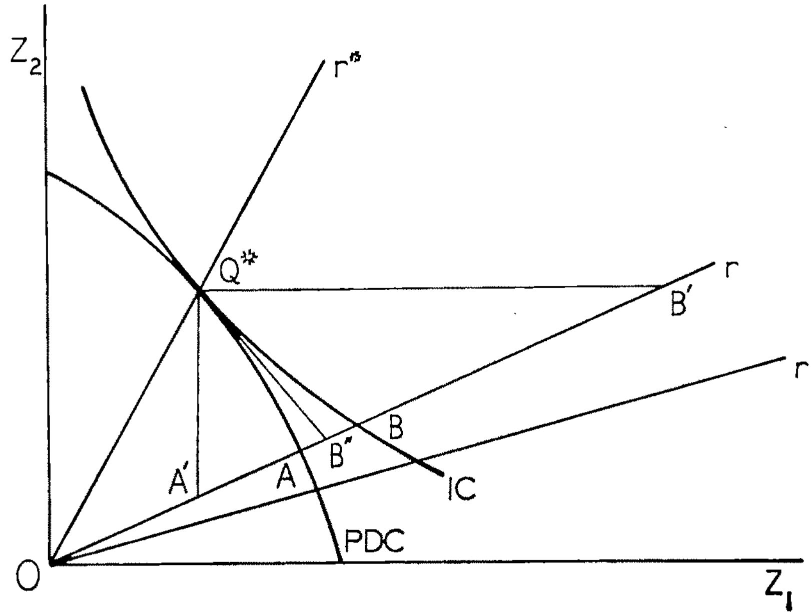


FIGURE 3

process to obtain some combination of the characteristics contents of the individual goods. Combinable consumption will be taken to be *linear*, the characteristics of the combination being the sum of the characteristics contents of the individual goods.[6] On the other hand, we may have *noncombinable* consumption in which only one good can be consumed at a time and characteristics can be obtained only in those proportions represented by an available good.

Both types of consumption are realistic in different contexts. Food nutrients fit the linear combinable case, but the services of many consumer durables fit the noncombinable pattern. The emphasis in this paper will be on noncombinable cases.

D. *Optimal and Sub-Optimal Transfer*

If we take a single individual in isolation and set out to minimize the resources required to attain a specified utility level, it is obvious which transfer good will achieve the optimum; it will be the good having a characteristics ratio r^* which corresponds to the point at which the relevant indifference curve for the consumer is tangent to a *PDC* (or at the corner of the indifference curve in the zero elasticity case). Figure 3 shows optimal transfer in diagrammatic form, and needs no further comment.

1. *Compensation for Sub-Optimal Transfer*

If the consumer's optimal good (that which would give him optimal transfer) is not available, there is some quantity of the "next best" good[7] that will enable him to achieve the utility level he would have attained with some specific quantity of the optimal good. If Q^* is the quantity of optimal good (characteristics ratio r^*) that enables the consumer to attain a specified welfare level, and Q is the quantity of a good with characteristics ratio r that enables him to achieve the same level, we can regard Q as the *compensating* quantity corresponding to Q^*. Since Q, Q^* are in comparable units, we can take the ratio Q/Q^* and refer to it as the *compensating ratio* for good r as compared with optimal good r^*.

---

[6] This is the consumption process on which my original analysis (1966) was based.

[7] It is assumed throughout the analysis that the substitution of characteristics from other product classes can be ignored, so that the "next best" good is always within the same product class as the "best" good.

Figure 3 shows a consumer for whom the optimal good is r^* and another good r ($r<r^*$) which is sub-optimal for this consumer. In order to bring the consumer to the same indifference level as attained with Q^*, he must be given an amount of good r which corresponds to point B on the diagram. From the properties of the quantity measure we are using, the compensating ratio is OB/OA.

It is obvious that the compensating ratio depends on three factors, the degree of curvature of the utility function, the degree of curvature of the *PDC*, and the difference between r and r^*. The ratio will increase with increased curvature of either curve and with an increase in the difference between r and r^*. The maximum possible value for the compensating ratio will be when B is at B' and A is at A' (corresponding to right-angled indifference and *PDC* curves with corners coincident at Q^*) and will be equal to r^*/r. In the case in which we have $r>r^*$, the maximum value of the compensating ratio will be r/r^*.

The compensating ratio will be one of the main tools of our subsequent analysis. Since it depends on r, r^* as well as on underlying preference and production conditions, it is useful to write the ratio as a function of r, r^*. We shall refer to the ratio in the form $h(r, r^*)$ as the *compensating function*, and it will have the following properties:

(1) $h(r^*, r^*)=1$

(2) $1 \leq h(r, r^*) \leq \text{Max}\,(r^*/r,\ r/r^*)$, for all r

(3) $h_r(r, r^*) \geq 0$ for $r \geq r^*$, ≤ 0 for $r \leq r^*$

(4) $h_{r^*}(r, r^*) \leq 0$ for $r \geq r^*$, ≥ 0 for $r \leq r^*$

(5) $h_{rr}(r, r^*) > 0$, unless both the indifference curve and *PDC* are linear,[8] at least for r within a "reasonable" distance of r^*.

II. Two Basic Theorems on Differentiation and Returns to Scale

Using the tools we have developed, it is possible to enunciate two basic theorems on the relationship between optimal product differentiation and returns to scale.

THEOREM 1: *If production is subject to constant or decreasing returns to scale, it will be socially optimal to produce every good which represents optimal transfer for any consumer. If and only if there are constant returns to scale in production, a linear product differentiation curve, and linearly combinable consumption processes, the social optimum can also be achieved with a number of goods in each product class not greater than the number of separate characteristics in the product class. Under decreasing returns to scale, it may be optimal to produce even more goods than under constant returns to scale with the same distribution of preferences, and certainly not less.*

Under constant returns to scale, the resources used to produce any quantity of any good are directly proportional to the quantity as measured, with the same constant of proportionality for all goods, because of the homotheticity assumption and the definition of the quantity measure. Any consumer who must consume a suboptimal good is therefore using more resources than would be needed for him to attain the same welfare level with an optimal good, and it is necessarily optimal to provide every consumer with his optimal good.

---

[8] This property may not hold for r sufficiently different from r^*, even in cases which would appear to be very well-behaved. For a linear *PDC* of the form $z_1+z_2=1$ and a Cobb-Douglas utility function of the form $u=z_1^{1/2}z_2^{1/2}$ (for which $r^*=1$), $h(r)$ has the form $h(r)=\frac{1}{2}(r^{1/2}+r^{-1/2})$, giving $h''=(3-r)/8r^{5/2}$. In this case $h''\geq 0$ only within the range $3r^*\geq r\geq 1/3r^*$.

If, in addition to constant returns to scale, there is linearly combinable consumption and a linear *PDC*, then production of the two extreme goods (each containing one of the characteristics only) will enable any consumer to attain the same welfare level by a combination of those goods as with his optimal good, for the same level of resource utilization.

Now consider the case in which there are decreasing returns to scale. In this case more goods, each produced in smaller quantities, will use less resources than fewer goods produced in larger quantities. It is obvious that it will never be optimal not to produce every consumer's optimal good. But now it may be optimal to have a consumer derive some of his welfare from a nonoptimal good, so as to reduce the average level of output of all goods.

Suppose that there is a linear *PDC* with equation $z_1+z_2=1$, and that there are two consumers whose preferences show zero elasticity of substitution and whose optimal goods are given by $r_1=3$, $r_2=1/3$. The input function has the form $v=Q^a$ ($a>1$ for decreasing returns to scale), and the consumers initially each receive one unit of their respective optimal goods for a total resource use of 2.

Now introduce a third good with characteristics ratio $r=1$, and suppose that each consumer receives some quantity of his optimal good, plus enough of the third good to bring him to his original welfare level. In this particular example the compensating ratios for the third good are equal at 1.5 and since the example is symmetrical we suppose each consumer receives an amount $(1-x)$ of his optimal good and an amount $1.5x$ of the third good, bringing each consumer to his original welfare level.[9]

Total resources used in the three-good case will then be given by

$$V = 2(1 - x)^a + (3x)^a$$

If we choose x so as to minimize V, we obtain

$$x^* = [1 + 3(3/2)^b]^{-1}$$

where $b=1/(a-1)$ which gives $x^*>0$ for all $a>1$, and thus it is optimal to produce some of the third good for all degrees of decreasing returns to scale. For $a=2$, we have $x=2/11$ and $V=18/11$, a saving of 4/11 resource units. Thus it will be optimal to produce goods which are not the optimal goods of any consumers, if there are decreasing returns to scale.[10]

It is interesting to consider the pricing associated with the above case. Since the compensating ratio is 1.5, a consumer will be induced to buy both the third good and his optimal good if the price of the third good is 2/3 that of the optimal good. Since the cost function for any good has the form $C=AQ^a$, where A is the same for all goods, we can compute the marginal costs of the goods at the optimum. It turns out that the ratio of the marginal cost of the third good to that of either optimal good is 2/3 for all values of a, when x is given its optimum value. Thus competitive prices are appropriate.

---

[9] Rather than having a single consumer with split consumption, which is somewhat unrealistic, we can suppose each "consumer" is actually an aggregate of consumers with identical tastes, some of whom consume one good and some the other.

[10] We have proved here only what we set out to do, showing that optimality under decreasing returns to scale will call for more goods than the number of distinct preferences. If decreasing returns holds for all levels of output, however small, we can continue to improve the situation by adding more and more goods (with characteristics ratios between those of existing goods) until we have a continuum of goods. If the preferences form a continuum, we will necessarily have a continuum of goods in the *constant* as well as decreasing returns cases.

A. *Increasing Returns to Scale*

Economic intuition suggests that the effect of increasing returns to scale will be to reduce the socially optimal number of goods below what it would be under constant returns to scale, at least for a sufficient degree of increasing returns. For a sufficiently well behaved case, the socially optimal number of goods will be a function of the degree of increasing returns and will decrease as this degree increases. Finally, intuition suggests that a solution for the optimal number of goods will necessarily involve solving for the optimal choice of those goods in terms of characteristics ratios. A grand theorem incorporating all these possible results is not easy to produce, so we shall start with a simple but basic theorem.

THEOREM 2: *There is some degree of increasing returns to scale, sufficiently large, for which the socially optimal number of goods is one.*

Consider a situation in which the consumers are supplied with N goods, the amount consumed of the ith good being Q_i. We make no assumptions about the specification of these goods or the distribution of preferences except that there are nonzero quantities of at least two goods. We shall consider the effect of replacing these goods by a single good, arbitrarily chosen with respect to characteristics. The quantity of the single good is denoted by Q and, if it is just sufficient to enable all consumers to attain the same welfare levels as with the N goods, it is obvious that $Q \geq \sum_1^N Q_i$. Write $Q = k\sum_1^N Q_i$, where $k \geq 1$.

We need first to prove that k is bounded above. Now, among all the consumers in the initial (N-good) situation, there must be one who consumes the greatest amount of characteristic z_1 per unit of whatever good he consumes; denote this maximal amount by $\bar{z}_1$. There will be a maximal z_2, denoted by $\bar{z}_2$, derived in a comparable fashion. Suppose the amounts of the two characteristics in unit quantity of the single good are z_1^*, z_2^* (note we restrict the single good to having some of both characteristics), then the maximal value for the compensating ratio cannot exceed $\text{Max}\,(\bar{z}_1/z_1^*, \bar{z}_2/z_2^*)$ for any consumer. The value of k cannot exceed the maximum compensating ratio for a single consumer, hence k is bounded.

Resource use in the N-good case is given by

$$V_N = \sum_1^N F(Q_i)$$

Since there are increasing returns to scale, $\sum F(Q_i) > F(\sum Q_i)$, with the ratio $\sum F(Q_i)/F(\sum Q_i)$ increasing as the degree of increasing returns increases.

If the N goods are replaced by a single good, produced in a quantity just sufficient to enable consumers to attain original welfare levels, resource use will be given by

$$V = F(Q) = F(k\sum Q_i)$$

With increasing returns, the ratio $F(k\sum Q_i)/F(\sum Q_i)$ is greater than unity, but approaches unity as the degree of increasing returns is increased without limit.

Thus we have the following situation. V_N is greater than $F(\sum Q_i)$ and recedes from $F(\sum Q_i)$ as the degree of increasing returns increases, while V is also greater than $F(\sum Q_i)$ but approaches it more and more closely as the degree of increasing returns increases. Thus there is some degree of increasing returns sufficiently great to give $V \leq V_N$, and the single good is then optimal for increasing returns of this or any greater degree, as compared with the specific N goods of the initial situation. Since no restrictions were placed on N or on the specification of the N goods, it follows that there is some degree of increasing returns for which the single good is optimal as compared with any choice of N goods, for all $N \geq 2$.

This theorem gives only a part of the overall picture with respect to increasing returns to scale, but firmly establishes the basic proposition that increasing returns will be associated with an optimal number of goods that is less than with constant returns to scale, and does this with the minimum of assumptions about preferences and their distribution or about the *PDC*.

III. The Conditions for Optimal Differentiation

Under constant returns to scale, the optimal number of product differentiates will be finite only if the number of distinct preferences represented in the society is finite, and is essentially unbounded if preferences form a continuum. Under decreasing returns to scale, the optimal number of goods will not generally be bounded even if the number of distinct preferences is finite. These are the conclusions to be drawn from Theorem 1. For increasing returns to scale, on the other hand, Theorem 2 shows the optimal number to be bounded for a sufficient degree of increasing returns. Thus the study of optimal differentiation is confined to the cases in which there are increasing returns.

Determining the optimal number of goods under increasing returns to scale is not a simple matter. We cannot merely consider the N-good case, then add an $(N+1)$th good, because there will be a total structural change in passing from N goods to $N+1$ goods. It is obvious that, since we are seeking optimality, we must find the specifications, quantities, and distribution of N goods which provide given welfare levels for minimum use of resources, then find specifications, quantities, and distribution of $N+1$ goods that achieve the same welfare levels as before for minimum resource use. In general, the goods in the $N+1$ case will be entirely different goods from those in the N-good case. The only thing that will be comparable between them will be the minimum resources required to attain the given welfare level in the two cases, and it is direct comparison of this that will indicate whether $N+1$ goods are better or worse than N. Thus solving for the optimal number of goods is a two-part process, first solving the optimum configuration for each possible number of goods, then comparing the resources required between different numbers of goods to find the optimum number. Since the number of goods is an integer variable and the configuration changes between different numbers of goods are discrete, the second stage is simple once the first has been completed.

A. *The Optimum Configuration*

The most difficult and interesting part of the overall problem is that of finding the optimum configuration for a specified number of goods. This involves finding characteristics ratios for the goods and the distribution of those goods over the consumers, such that the given welfare levels for all consumers are achieved with minimum use of resources. It is obvious that opportunities for ill behavior abound in a general model, so we shall make the two following assumptions on the distribution of preferences which are designed to give a minimum level of good behavior:

Continuum Assumption. The distribution of preferences is such that the set of all points along the *PDC* that represent optimal transfer for some individual forms a continuum.

Anticrossover Assumption. The distribution and form of preferences is such that, if it is optimal to supply two consumers whose optimal transfer goods would have characteristics ratios r_i, $r_j (\leq r_i)$ with the same good Q_k, then it will be optimal to supply Q_k to all consumers whose optimal good would have a characteristics ratio r such that $r_j \leq r \leq r_i$.

The two assumptions between them ensure that the set of all consumers being supplied with the same good is compact. It is possible to visualize cases, such as two consumers with the same optimal good, one with zero elasticity of substitution and the other with infinite elasticity, in which it just might sometimes be optimal to supply them with different goods. We rule this out by the anticrossover assumption, which implies that the shape of preferences as between individuals with closely similar optimal transfer goods does not vary too much.

To determine the optimal configuration for N goods we proceed as follows. First we note that, as a consequence of the above assumptions, the continuum of consumers (identified by the characteristics ratios of their optimal transfer goods) will be divided into N segments, each segment supplied by one of the N goods. Denote by $R_1, \ldots, R_{N-1}$ the characteristics ratios which divide the segments from each other, and by $r_1, \ldots, r_N$ the characteristics ratios of the N goods. For given preference and population distributions and given production conditions, the quantity of the ith good (characteristic ratio r_i) required to bring all consumers in the segment bounded by R_i, R_{i-1} to specified welfare levels will be a function of r_i, R_i, R_{i-1} only, so that we can write Q_i as $Q_i(r_i, R_i, R_{i-1})$.

Let us take R as given, and consider the optimal choice for r_i. Since Q_i is not a function of r_j, nor Q_j of r_i, and since R is held constant, Q_i is independent of Q_j (all i, j) and thus the optimal choice for r_i must be that which minimizes Q_i. This gives us our first optimum condition

$$(6) \qquad \partial Q_i(r_i, R_i, R_{i-1})/\partial r_i = 0$$

If Q_i^* is Q_i optimized with respect to r_i, then $Q_i^* = Q_i^*(R_i, R_{i-1})$. We can now find the optimal value for R. The Q^*, considered as functions of R, are no longer independent of each other because each R_i appears as an argument in two of the Q^*s. To optimize we must minimize total resources, given by

$$V = \sum F_i(Q_i^*) = V(R_1, \ldots, R_{N-1})$$

Because R_j appears as an argument in Q_j^*, Q_{j+1}^* only, the optimum conditions for R takes the form

$$(7) \qquad \frac{\partial V}{\partial R_i} = \frac{\partial Q_{i+1}}{\partial R_i} F'_{i+1} + \frac{\partial Q_i}{\partial R_i} F'_i = 0$$

B. *Interpretation of the Optimum Conditions*

We can give the optimum conditions an interpretation with more direct economic appeal by expressing them in terms which involve the compensating ratios. To do this we shall make the additional assumption, that *all consumers having the same optimal transfer good have identical preferences.*

Consider the segment of consumers bounded by optimal transfer ratios R_i, R_{i-1}, all supplied by good G_i with characteristics ratio r_i. Consider those consumers whose optimal good would have characteristics ratio r $(R_i \geq r \geq R_{i-1})$, and let the quantity of that good that would bring all these consumers to the specified welfare levels be denoted by $s(r)$. Since the consumers are being supplied with a good having characteristics ratio r_i, not necessarily their optimal good, the total quantity of this good needed to achieve the welfare level is given by $h(r_i, r)s(r)$, where h is the compensating function as discussed in Section I. The total quantity of good G_i required to bring all consumers in this segment up to specified utility levels is then given by

$$Q_i = \int_{R_{i-1}}^{R_i} h(r_i, r)s(r)dr$$

Since R_i, R_{i-1} appear only as limits of integration, we thus have

$$\partial Q_i/\partial R_i = h(r_i, R_i)s(R_i)$$
$$\partial Q_i/\partial R_{i-1} = -h(r_i, R_{i-1})s(R_{i-1})$$

Inserting these values in the optimum conditions (7) gives

$$\frac{F'_i}{F'_{i+1}} = \frac{h(r_{i+1}, R_i)}{h(r_i, R_i)} \tag{8}$$

Now F'_i, F'_{i+1} are the marginal resource costs of producing goods G_i, G_{i+1}, respectively. The right-hand side of (8) is the ratio of the compensating ratios for the dividing consumer (whose optimal transfer good is on the boundary between the two adjacent segments) with respect to being supplied with G_{i+1} as compared with G_i.

Thus the second optimum condition, in the case in which consumers having identical optimal transfer goods have identical preferences, has the following easily interpretable form: *The ratio of the marginal resource costs of two adjacent goods will be the inverse of the ratio of the compensating ratios with respect to the two goods, for the dividing consumer.* No special insights are given by this formulation into optimum condition (6) or into the optimality of N goods relative to M goods.

C. *A Well-Behaved Example*

To show that intuitive notions about well-behaved relationships between the optimal number of goods and the degree of returns to scale are valid in appropriate cases, we shall sketch out and give the results for a simple example.

On the production side we assume a linear *PDC*[11] of the form $z_1 + z_2 = 1$ and an input function of the form $V = Q^{1/b}$ where the degree of increasing returns increases with b. We assume that consumers show zero elasticity of substitution, have optimal goods distributed with uniform line density along the *PDC*,[12] are to be maintained on welfare levels corresponding to unit quantities of their optimal goods, and that the total population is unity.

It can be shown that, under these assumptions, the compensating function h and the density function s have the following forms:

$$\begin{aligned} h(r_i, r) &= \frac{r(1+r_i)}{r_i(1+r)} & r &\geq r_i \\ &= \frac{1+r_i}{1+r} & r &\leq r_i \\ s(r) &= \frac{1}{(1+r)^2} \end{aligned}$$

where r is the characteristics ratio of the optimal good and r_i the characteristics ratio of the good actually supplied to consumers with optimal good r.

Confining our investigations to the cases $N = 3, 2, 1$, we obtain the following optimal configurations for the three cases:

$$\begin{array}{llll} N=3: & r_1 = 1/3, & Q_1 = 7/18 & \\ & r_2 = 1, & Q_2 = 7/18 & R_1 = 1/2, R_2 = 2 \\ & r_3 = 3, & Q_3 = 7/18 & \\ N=2: & r_1 = 1/2, & Q_1 = 5/8 & \\ & r_2 = 2, & Q_2 = 5/8 & R_1 = 1 \\ N=1: & r_1 = 1, & Q_1 = 3/2 & \end{array}$$

Denoting by $V(N)$ the resources required to achieve the given utility levels with N goods of optimal configuration, we obtain $V(3) = 3(7/18)^{1/b}$, $V(2) = 2(5/8)^{1/b}$, $V(1) = (3/2)^{1/b}$. By solving for the values of b at which the various pairs are equal, we obtain our final result: The optimal number of goods is three for $b \leq 1.169$ (relative to the choice between 3 and a lesser number of goods), two for $1.169 \leq b \leq 1.263$, and one for $b \geq 1.263$. Note that the total quantity of goods ($\sum Q_i$) increases from 7/6 with three goods to 5/4 with two goods to 3/2 with one good (it is 1 for $b=1$, the constant returns to scale case). If this were not so, the optimal number of goods would never be greater than unity, if there were any degree of increasing returns.

---

[11] With a linear *PDC* we shall assume consumption is noncombinable, to rule out the possibility of achieving optimality with no more than two goods.

[12] Uniform line density is one of several possible cases of "uniform" preference distribution.

This example shows that, for a sufficiently well-behaved model, we do find the optimal number of goods decreasing steadily as the degree of increasing returns increases.[13]

IV. Four Theorems on Pricing and Imperfect Competition

The conditions for optimal configuration at the optimum number of goods lead directly to the following theorem:

THEOREM 3: *If the optimal number of goods is some number $N > 1$, then the optimal prices for those goods will be such that prices stand in the same ratio to marginal costs for all goods in the group. This can be shown to hold for any PDC, any forms of the preference functions,*[14] *and any forms for the input functions, provided there is a proper interior optimum at N goods.*

Consider the optimal distribution of the goods through a market mechanism (where each consumer has an income appropriate to the specified welfare levels), and consider the consumer whose optimal transfer good would have characteristics ratio R_i, the dividing point between the market segments supplied by G_i and G_{i+1}. If he is to be the dividing consumer, he must find that the dollar expenditures on G_i or G_{i+1} which are equivalent to a unit of his (unavailable) optimal good are the same. Thus the ratio P_i/P_{i+1} must be inverse to the ratio of the two compensating ratios $h(r_i, R_i)$, $h(r_{i+1}, R_i)$ for the dividing consumer. But we have already shown in equation (8) that the ratio of marginal costs, F_i'/F_{i+1}', must also equal the inverse of the ratio of the compensating ratios at the optimum configuration for N goods. Thus the ratio of prices between adjacent goods must equal the ratio of their marginal costs. By chain reasoning, the same relationship must then hold for all pairs of goods, proving the theorem.

Note that the theorem requires only that all prices bear the same ratio to marginal costs for goods within the group, and does not require equality of prices and marginal costs. However if we consider the group embedded in the larger economy then the usual arguments will lead to the requirement of equality in order to achieve optimality over the economy as a whole.

Equalization of the ratios of prices to marginal costs within the group is necessary but not sufficient for optimal configuration, and certainly not sufficient to guarantee that the number of goods is optimal (as we shall see in Theorem 6). It is, of course, easily possible to have the economy operating in an optimal configuration for N goods but be sub-optimal because the optimal ***number*** of goods is really M.

A. *Achieving the Optimum*

Since the interesting problems with respect to the optimum number of goods

---

[13] *A Conjecture*: The pattern of values for $\sum Q_i$ suggests that for N goods we would have $\sum^N Q_i = (2N+1)/2N$ with the optimal Q_i equal to each other at a value of $(2N+1)/2N^2$. If the conjecture were indeed true, we could then show that the value of b at which society would be indifferent between N goods and $N+1$ goods would be given by

$$b = 1 + \frac{\log\,[1 + 1/N(2N+3)]}{\log\,[1 + 1/N]}$$

This would give the lowest value of b for which the optimal number of goods would be three as 1.128 (the upper value for three goods is 1.169). The range of values for which the optimal number of goods would be ten would be $1.046 \leq b \leq 1.051$. I have not attempted to verify the truth of the conjecture.

[14] As proved, we require that all consumers with the same optimal goods have the same preferences. This can probably be relaxed.

arise with increasing returns to scale, and since the optimum pricing system is that of perfect competition, a problem arises in achieving that optimum. The competitive system will not work, first because obtaining the potential scale economies for each good requires a single producer for that good (unless the economies are all industry externalities) and, second, because all firms will make losses when price is equated to marginal cost. Furthermore, even marginal cost pricing cannot guarantee that the *number* of goods is optimal.

One possibility is a managed economy with single firm control of the production of each good and each firm constrained to adopt marginal cost pricing. This would then require a subsidy to cover the gap between total costs and total revenue for each firm. The subsidy itself would be a control variable however, since the optimum number of goods would require the minimum subsidy over the group. A subsidy set at this minimum level for the group as a whole should then induce the group to produce the optimum number of goods, given appropriate institutional rules.

B. *Imperfect Competition*

The remainder of the paper is concerned with imperfect competition and its effects on the optimal number, quantity, and specification of goods. We have different cases to consider, but in each case we shall commence from the optimum position and consider the changes introduced by the relevant type of imperfect competition. First we shall clear the air by disposing of a simple but important case.

THEOREM 4 (Constant Returns to Scale): *Under constant returns to scale, market imperfection will not cause the number or specification of goods to diverge from the optimum.*

If we commence at the optimum and there are constant returns to scale, no form of market imperfection will make it profitable for any firm to eliminate anyone's optimal transfer good because 1) there is no saving in costs from changing the number of goods produced, and 2) the maximum revenue that can be extracted from any one consumer for a given quantity of good supplied will be when he is sold his optimal transfer good. Thus the same goods will be sold, with the same characteristics ratios, as at the optimum. If there are any gains to be made by the monopolist, they will be by increasing prices and reducing quantities of the same goods which are produced at the optimum, not by changing the number of goods or their characteristics ratios. The remaining theorems are all presumed to refer to a context of increasing returns to scale.

THEOREM 5 (Single Good Monopoly): *Monopoly control of the production of any one good when firms producing other goods do not behave as monopolists will in general lead to a nonoptimal choice for the characteristics ratio of that good and, if other firms adopt marginal cost pricing, to a restriction of the output of the monopoly good.*

The second part of the theorem should surprise no one, but the first part represents a new kind of result for imperfect competition theory.

To prove the theorem, consider a particular good (characteristics ratio r) in the overall spectrum. Next to this good in the spectrum will be an "upper" good (characteristics ratio $r_U > r$) and a "lower" adjacent good (characteristics ratio $r_L < r$). The dividing consumers between the markets for the good in question and the upper and lower adjacent goods have optimal transfer goods with characteristics ratios R_U, R_L. We start from the optimal configuration and consider the changes that will be made by a monopolist taking control of the good.

We assume that over the range of variations being considered, there are no acceptable substitutes for the good from outside the group. If the price of the good increases, consumers in the center of the market will simply have to buy the same quantity of the good and spend more in order to remain at the same welfare level. Consumers at the market fringes, however, may switch to the purchase of adjacent goods if it is advantageous to do so. The market boundaries will adjust so that the dividing consumer is the one for whom the ratio of compensating ratios is inversely proportional to the price ratio. Thus the quantity of the good will be given by

$$(9) \qquad Q = \int_{R_L}^{R_U} h(r, x)s(x)dx$$
$$= Q(r, R_U, R_L)$$

where $h(r, x)$ is the compensating ratio for the consumer whose optimal good has characteristics ratio x, and $s(x)$ is the market density at x. The effect of price changes operates through changes in R_U, R_L.

By definition of the dividing ratios, the following conditions must be satisfied everywhere:

$$(10) \qquad h(r, R_U)P = h(r_U, R_U)P_U$$

$$(11) \qquad h(r, R_L)P = h(r_L, R_L)P_L$$

where P, P_U, P_L are the prices of the good in question and the upper and lower adjacent goods, respectively.

Relationships (9), (10), (11), together with the profit definition

$$(12) \qquad \pi = PQ - C(Q)$$

(where $C(Q)$ is the cost function) form a set of four relationships that must be satisfied everywhere by the six variables π, P, Q, r, R_U, R_L. There are two degrees of freedom in the system, which we shall take to be the choice of r and the choice of Q.

To prove the first part of the theorem, hold Q constant and consider the effect of variations in r on the firm's profit. Using standard comparative static methods and noting that $\partial Q/\partial r = 0$ at the optimum (optimum condition (6)), we obtain

$$(13) \qquad \frac{d\pi}{dr} = \frac{Q_L M_U h_r^L + Q_U M_L h_r^U}{Q_U M_L h^U + Q_L M_U h^L} PQ$$

at the optimum point, where

$$h^U = h(r, R_U), \qquad h^L = h(r, R_L)$$
$$h_r^U = \partial h^U/\partial r, \qquad h_r^L = \partial h^L/\partial r$$
$$M_U = \partial[Ph^U - P_U h(r_U, R_U)]/\partial R_U$$
$$M_L = \partial[Ph^L - P_L h(r_L, R_L)]/\partial R_L$$
$$Q_U = \partial Q/\partial R_U = h^U s(R_U)$$
$$Q_L = \partial Q/\partial R_L = -h^L s(R_L)$$

The h-functions always decrease as the two arguments move closer together and always increase as they diverge, so that M_U, h_r^L are positive and M_L, h_r^U are negative, while h^U, h^L are essentially nonnegative; Q_U is positive, and Q_L negative. Thus the denominator in (13) is always negative, but the numerator consists of two terms of opposite sign. Thus $d\pi/dr$ may have any sign (or be zero but only by coincidence), and the monopolist will in general find it profitable to change the characteristics ratio of his good away from the optimum.

The reason why the monopolist will not generally be satisfied with the socially optimum specification of the good is not difficult to see. From his point of view, the central consumers are locked into his market and it is only at the edges where he must compete with adjacent goods. Thus his actions in the market will be based on the properties at the market fringes, as shown by (13) where $d\pi/dr$ is seen to depend only on the properties at R_U, R_L. The socially optimal choice of r, on the other hand, is given by:

$$\frac{\partial Q}{\partial r} = \int_{R_L}^{R_U} \frac{\partial h(r, x)}{\partial r} s(x)dx = 0$$

a condition which gives weight to all consumers in the market segment.

To prove the second part of the theorem, we hold r constant and consider variations in Q. Proceeding as before, we obtain

$$(14) \qquad \frac{d\pi}{dQ} = (P - C') - \frac{QM_U M_L}{D}$$

where D is the same as the denominator in (13).

If, as assumed, we have marginal cost pricing at the optimum, then $P-C'=0$ and we have

$$(15) \qquad \frac{d\pi}{dQ} = -\frac{QM_U M_L}{D}$$

Since M_U, M_L have opposite signs and D has already been shown to be negative, $d\pi/dQ<0$ and the monopolist will find it profitable to reduce output (necessarily involving an increase in price and a shrinking of the market boundaries), completing proof of the theorem.

The next theorem is similar in its policy conclusions to the well-known "excess capacity" theorem of monopolistic competition, although it is based on a different process of reasoning and does not depend in any way on the ambiguous notion of "capacity."

THEOREM 6 (Monopolistic Competition): *Under increasing returns to scale and with a sufficiently uniform and evenly distributed market, "monopolistic competition" will lead to a greater degree of product differentiation than is socially optimal.*

Quotation marks have been placed around monopolistic competition here because one of the most basic assumptions of the Chamberlin model, that the effects of a behavior change by any one firm will be spread evenly over all other firms, most emphatically does not hold in this case. Here, the behavior of a firm affects very much those firms producing adjacent goods in the spectrum, and very little those producing goods remote from it in the spectrum. We make the remaining assumptions of monopolistic competition, that each firm produces one product, no other firm produces the same product, and there is free entry into the group that will continue so long as positive profits can be made. Thus, although the dynamics of the group cannot conform to the Chamberlin model because of the oligopolistic elements present, the equilibrium (if attained) will be of the Chamberlin kind—for a sufficiently evenly distributed total market, all firms will be of the same size, with marginal revenues equal to marginal costs and prices equal to average costs. Thus we shall consider the term *monopolistic competition* to be the most appropriate for the model being considered.

A "uniform" market is taken here to mean that, after a possible transformation of coordinates, the compensating ratio for a consumer whose optimal good has characteristics ratio corresponding to x, with respect to a good having characteristics ratio corresponding to x^*, depends only on the absolute value of the difference between x and x^*. That is, the compensating function has the particular form $h(x^*, x) = h(|x-x^*|)$. If the upper and lower dividing consumers for the market for the good represented by x^* have optimal goods with characteristics ratios x^*+c_2, x^*-c_1, and the market density is constant (taken to be unity), total demand for the good represented by x^* is given from:

$$(16) \qquad Q = \int_{x^*-c_1}^{x^*+c_2} h(|x - x^*|)dx$$
$$= H(c_1) + H(c_2)$$

where $dH/dx = h$.

Given this uniformity and uniform

density, it is obvious that goods and firms will be arranged, both at the social optimum and at monopolistic competition equilibrium, so that $c_1=c_2$ for every good, and c and Q are equal for all goods. Thus we can write $h(c)=H'(c)$ and $Q=2H(c)$ for all goods.

Since the whole spectrum is to be covered by goods with equal market areas (in the coordinates being used), there is an inverse relationship between the number of goods n, and the market area measure c, which we can write as $n=A/c$ where A is an appropriate constant.

Now consider the optimum number of goods, and suppose it is sufficiently large to treat n as closely approximated by a continuous variable. For n goods, total resource use is given by

$$V = nF(Q) \tag{17}$$

where $F(Q)$ is the input function for the single good. Since both n and Q are functions of c only, we minimize V with respect to c to obtain the condition for the optimum

$$F' = F/2ch \tag{18}$$

where the second-order conditions are satisfied for increasing returns to scale.

Having established the optimum conditions, we turn to consider the monopolistic competition equilibrium, which will consist of firms facing identical conditions. Each firm will, since we assume no oligopolistic elements and thus that other prices are taken as given, face the same considerations as the firm of Theorem 5. Due to the special uniformity assumptions, the expressions that appear in Theorem 5 have the following simplified forms here

$$h^U = h^L = h \quad (=h(c))$$

$$M_U = - M_L = 2Ph'$$

$$Q_U = - Q_L = h$$

Inserting these values in (14) of Theorem 5, together with the value of $Q(=2H)$ from above, and noting that $h=H'$, $h'=H''$, we obtain

$$\frac{d\pi}{dQ} = P - C' - 2P\frac{HH''}{(H')^2} \tag{19}$$

We shall suppose prices to be measured in resource units, so that $C=F$, $C'=F'$. We commence at the optimum and thus assume that the condition (18), $F'=F/2cH'$ is satisfied.

There are two equilibrium conditions for the monopolistic competition group: 1) that all firms are at individual profit maximizing equilibria; and 2) that there are just sufficient firms to equate price and average cost everywhere. We shall proceed by supposing that we have the socially optimal number of goods and that price is equal to average cost for all firms, then investigate whether the firms are at profit maximizing equilibria.

Inserting the values $C'=F'=F/2cH'$, $P=F/2H$, in (19) we obtain, after some manipulation,

$$\frac{d\pi}{dQ} = \frac{F}{2H'}\left[\frac{H'}{H} - \frac{H''}{H'} - \frac{1}{c}\right] - \frac{FH''}{2(H')^2} \tag{20}$$

Since H'' is positive (the compensating ratio is an increasing function of c), the last term above is clearly negative. Now consider the expression in brackets. This is equal to $d(log\ G)/dc=G'G$, where $G=H/cH'$ and is thus equal to the ratio of the average to the marginal compensating ratio over the half market from $x=x^*$ to $x=x^*+c$. Due to the assumed shape of the compensating function, G is a decreasing function of c so that G' is negative, the bracketed expression is negative, and $d\pi/dQ$ is negative.

Thus a configuration with the optimal number of goods and zero profits will not be profit maximizing for the individual firms. These can increase their profits by contracting output, leading to positive profits and the entry of more firms. The final monopolistic competition equilibrium will therefore result in a greater number of goods than at the social optimum.

Finally we shall consider the effect of monopolization of a market sector, that is control of the outputs of several goods (taken to be adjacent in the spectrum) by a single monopolist.

THEOREM 7 (Monopoly Control of a Market Sector): *Under increasing returns to scale, monopoly control of a market sector will lead to a lesser degree of product differentiation over that sector than is socially optimal.*

Consider an optimal configuration over an optimal number of goods, and consider the effect of monopoly control over several adjacent goods in the market. We shall suppose that the monopolist takes over control of the $(j-1)$th, jth and $(j+1)$th goods. Although not essential, it simplifies the argument to suppose that there is uniform distribution over this sector so that, at the social optimum, the prices of all three goods are the same.

Consider the effect of taking the jth good out of production. This will not affect those consumers who originally purchased the other two goods and, if prices are unchanged, will have no effect on consumers purchasing goods outside the monopolized sector. Since we assume throughout that there are no acceptably close substitutes for goods within the general product class from outside the class, those consumers who originally purchased the jth good will now purchase either the $(j-1)$th or the $(j+1)$th. If their incomes are fixed, they will obtain the same amount and the same quantity of the substitute as of their original good (since prices are the same), but will be somewhat worse off. If they are compensated, they will buy more of the substitute than of the original good. Thus the monopolist's average revenue will be unchanged on total sales no less than with all three goods.

Since there are increasing returns to scale, however, the increased output of the $(j-1)$th and $(j+1)$th goods will reduce average costs and thus the monopolist's profits will be increased as a result of closing down production of the jth good. Simply closing down the production of the jth good does not represent the monopolist's final equilibrium position, of course, but the profitability of this move is sufficient to prove the theorem.

V. Policy Conclusions

The policy conclusions to be drawn from the analysis can best be put in the form of answers to five questions which the policy maker must inevitably ask.

1. *Is there a socially optimal degree of product differentiation?* The answer to this basic question is yes, there is a socially optimal degree of product differentiation, divergence from which will increase the resources needed to enable consumers to attain specified levels of welfare.

2. *Can the optimum, or divergences from it, be easily recognized in the economy or in a particular industry?* The answer to this question seems, alas, to be no. Although there are clearly defined conditions that must be satisfied for the optimal configuration for a specific number of product differentiates, there are no easily recognizable conditions with respect to the actual number of goods.

3. *Would perfect competition throughout the economy result in attainment of the optimum?* The pricing system appropriate to the optimal configuration would be established under perfect competition. However, the problem of optimal differentiation is most important and most interesting under increasing returns to scale. Perfect competition, under these circum-

stances, could not take advantage of the scale economies and thus would not generate the optimum, which would require marginal cost pricing and single firm output for each good. Thus we cannot look to perfect competition to solve the optimum differentiation problem.

4. *Will market imperfections tend to give a nonoptimal degree of product differentiation?* Under constant returns to scale, no. Under increasing returns to scale, yes.

5. *Does market imperfection give a consistent bias in the degree of product differentiation, always too little or always too much?* Under increasing returns to scale, when imperfect competition tends to give nonoptimal differentiation, the direction of bias depends on the exact market structure. Monopolistic competition will lead to too much differentiation, monopolization of a market sector to too little. More complex, and thus more realistic, market structures may be expected to show effects of both kinds, leaving the direction of bias uncertain.

REFERENCES

H. Hotelling, "Stability in Competition," *Econ. J.*, 1929, *39*, 41–57.

K. Lancaster, "A New Approach to Consumer Theory," *J. Polit. Econ.*, Apr. 1966, *74*, 132–57.

———, *Consumer Demand: A New Approach*, New York 1971.

Security Price Changes and Transaction Volumes: Theory and Evidence

By THOMAS W. EPPS*

A familiar Wall Street adage is that transaction volume (the number of shares traded) tends to be relatively high in bull markets and low in bear markets. Although economists have long been interested in the behavior of speculative prices, they have not given adequate attention to the relation between transaction volume and price change.

Charles Ying examined the relation between the Standard and Poor's index of daily closing prices of 500 common stocks and total daily volume on the New York Stock Exchange (*NYSE*). He found that large increases in volume are usually accompanied by large price changes, that large volumes are typically associated with increases in price, and that small volumes usually accompany price declines. Unfortunately, the validity of Ying's findings has been questioned because of his adjustments on the data to account for the effects of dividends and new issues of stock over the sample period. More fundamentally, the two series that he examined (the 500-stock index and total *NYSE* volume) are not comparable.

The second criticism also applies to tests by C. W. J. Granger and Oskar Morgenstern, who found that cross-spectral analysis of weekly data showed no evidence of a relationship between the Securities Exchange Commission (*SEC*) composite price index and total *NYSE* volume. In a later paper, however, Michael Godfrey, Granger, and Morgenstern drew the same conclusion from tests on price and volume for two stocks, using individual transactions data. But none of these tests was designed to detect a relation between the volume on a transaction and the *change* in price from the previous transaction, or even between total volume and price change over an interval of time.

Their failure to detect significant correspondence between price and volume led Godfrey, Granger, and Morgenstern to question the applicability of existing theory. In their words:

> A basic problem which this analysis points out clearly is the lack of a useful theoretical framework in economics for the analysis of speculative markets . . . . It is clear that in the situation described here the volume of shares transacted bears no discernable relationship to any conventional concept of supply or demand. [p. 24]

In this paper I will show that such a theoretical framework can be constructed from a fairly broad set of portfolio-selection models and that the resulting theory implies a relationship between volume and price change on individual transactions which is in accord with the market adage stated in the opening paragraph. Specifically, the theory will be shown to imply that the ratio of volume to price change for upticks exceeds the absolute value of this ratio on downticks. The model from which this hypothesis is derived serves to bridge somewhat the gap between portfolio-selection theory and the theory of speculative prices.

---

* Assistant professor, University of Virginia. I wish to express appreciation to Mary Lee Epps for writing the computer programs used in this study and to Roger Sherman for comments on an early draft. I am especially grateful to a referee for several important suggestions and to the Wilson Gee Institute for financial assistance.

Section I contains a development of the model; Section II, the empirical test; and Section III, a summary and conclusions.

I. Theory

The development of the theory is in two parts. Using a "two-parameter" portfolio-selection model as a framework, I show in Sections IA and IB that the magnitude of the intercept and slope of an individual's demand function for an asset increase with the asset's expected end-of-period value. Section IC examines the implications of this fact for market prices and transaction volumes.

A. *The Individual's Demand Function*

The following assumptions and conventions will be used:

1. Individuals base their choices among portfolios on the expected values and variances of their end-of-period values (*EPV*). The "end of the period" is some fixed moment in time.
2. An individual's utility increases monotonically with the expected *EPV* of his portfolio and decreases monotonically with its variance (risk), but the extent of his risk aversion (defined below by equation (8)) does not change with expectation and variance of *EPV*.
3. There are $n+1$ assets. Everyone considers the first n to be risky and the last one to be riskless. Strictly to simplify the exposition, we regard the riskless asset as yielding no return. (No essential property of the model is altered by this convention.)
4. There are no restraints on borrowing or selling short: individuals may freely trade claims against themselves for units of any of the $n+1$ assets.[1]
5. Anticipations about *EPV*s of assets have zero elasticity with respect to current prices.
6. Each individual considers the *EPV* of each (risky) asset to be a random variable, whose distribution changes in response to news shocks in such a way that coefficient of variation remains constant.

The last assumption requires explanation. Let $\tilde{p}_k$ be the *EPV* of the kth asset, and let p_k and S_k^2 be its expected value and variance, respectively, in the opinion of some investor. The assumption states that news which alters S_k and p_k leaves the ratio S_k/p_k unchanged. This would occur, for example, if $\tilde{p}_k$ were lognormally distributed and if news affected the mean (μ_k) but not the variance (σ_k^2) of the subjective distribution of *log* $\tilde{p}_k$. The last statement is true because lognormality implies[2]

$$p_k = \exp\left(\mu_k + \frac{\sigma_k^2}{2}\right) \tag{1}$$

$$S_k^2 = p_k^2(\exp \sigma_k^2 - 1) \tag{2}$$

so that

$$S_k/p_k = (\exp \sigma_k^2 - 1)^{1/2} \tag{3}$$

The lognormality hypothesis itself is quite plausible. Early investigators[3] of the distribution of stock prices argued that changes in the logarithm of price over intervals of time ($\Delta\, log\, P_t$) are normally distributed. While subsequent research has shown that infinite-variance distributions of the stable class and finite-variance non-Gaussian mixtures of normals usually better characterize $\Delta\, log\, P_t$, normality remains a useful hypothesis in some applications. If the distribution of $\Delta\, log\, P_t$ is regarded as approximately normal, it is reasonable to suppose that investors infer from past price behavior that changes in the logarithm of price over future periods are normally distributed as well. This implies that they regard end-of-period price as lognormal. Furthermore, if holding periods are short (so that anticipated dividend or interest payments are small relative to price), then *EPV* is approximately lognormal.

---

[1] Borrowing of cash or other assets is regarded as being accomplished in the following manner. The borrower executes an agreement with the lender to return the borrowed units of the asset at the end of the period, together with any emoluments (such as dividends) that the owner would normally receive. The borrower then trades the borrowed units for units of more preferred assets. Naturally, no new units of the asset are created on balance by the transaction—the holdings of the lender, the borrower, and the persons with whom the borrower has traded summing to the original quantity. Borrowing agreements are considered to be renegotiated at the end of each period; and periods are short enough that we may regard as negligible the risk of sufficient price fluctuation to cause default on the agreements.

[2] See J. Aitchison and J. A. C. Brown.

[3] See, for example, the papers of M. F. M. Osborne and of Arnold Moore. A reasonably complete bibliography of more recent work is given in the paper by the author and M. Epps.

B. *A Model of Investor Behavior*

The process of portfolio choice may be characterized as the search for the set of asset quantities, $a_1, a_2, \ldots a_n, a_{n+1}$, which maximizes

$$U = U(a'p + a_{n+1}, a'Sa) \tag{4}$$

subject to $W = a'P + a_{n+1}$

where 1) a, p, and P are, respectively, column vectors of the quantities, the expected *EPV*s, and the market prices of the n risky assets; 2) S is a (nonsingular) variance-covariance matrix of *EPV*s of the risky assets; and 3) W is the individual's wealth, or net worth. The first argument of U is the expected *EPV* of the portfolio; the second argument is the variance of the *EPV* (risk). Assumption 2 requires that $U_1 > 0$ and $U_2 < 0$. Assumption 3 implies that the expected *EPV* and price of the riskless asset ($n+1$) are equal, and we have arbitrarily set $P_{n+1} = 1$.[4]

The first-order conditions for a constrained maximum of U are

$$U_1 p_i + 2U_2 \sum_{m=1}^{n} a_m S_{im} - \lambda P_i = 0, \qquad i = 1, 2, \ldots, n \tag{5a}$$

$$U_1 - \lambda = 0 \tag{5b}$$

$$W - a'P - a_{n+1} = 0 \tag{6}$$

where S_{im} is the covariance between *EPV*s of the ith and mth assets. Equation (5b) gives $\lambda = U_1$. Dividing both sides of the first n equations by U_1 and solving for a yields

$$a = (\beta S)^{-1} \cdot (p - P) \tag{7}$$

where

$$\beta \equiv -2U_2/U_1 \tag{8}$$

A positive scalar, β represents the coefficient of risk aversion.

Since β is constant (by assumption 2), equation (7) provides solutions for the optimal quantities of all the risky assets. These are determined independently of the wealth constraint (6), which serves merely to specify the (positive or negative) value that must be assumed by a_{n+1}. For the kth risky asset we have

$$a_k = \frac{1}{\beta} \sum_{i=1}^{n} \frac{s_{ik}}{s} (p_i - P_i) \tag{9}$$

where s_{ik} is the cofactor of S_{ik} in S, and where s is the determinant of S. Now

$$s = r \prod_{i=1}^{n} S_i^2 \tag{10}$$

$$\frac{s_{ik}}{s} = \frac{r_{ik}}{r S_i S_k} \tag{11}$$

---

[4] The model is a fairly typical two-parameter model, resembling those of William Sharpe, John Lintner, Jan Mossin, G. O. Bierwag and M. A. Grove, and Donald Farrar. The Sharpe-Lintner models have investors maximizing preference functions of the expected value and variance (or standard deviation) of *rate* of return subject to a market constraint of attainable return-risk combinations; but despite their apparent dissimilarity to the model described here, they can be shown to be equivalent in essential respects.

where r is the determinant of the correlations matrix R, derived from S, and where r_{ik} is the cofactor of element R_{ik}. It follows from equations (9) and (11) that

$$a_k = \frac{1}{\beta} \sum_{i=1}^{n} \frac{r_{ik}(p_i - P_i)}{r S_i S_k} \tag{12}$$

The demand function specified by equation (12) may be expressed in Marshallian form:

$$P_k = p_k + S_k \sum_{i \neq k} \frac{r_{ik}(p_i - P_i)}{r_{kk} S_i} - S_k^2 \beta \frac{r}{r_{kk}} a_k \tag{13}$$

Let

$$\alpha_k \equiv 1 + \frac{S_k}{p_k} \sum_{i \neq k} \frac{r_{ik}(p_i - P_i)}{r_{kk} S_i} \tag{14}$$

$$\gamma_k \equiv S_k^2 \beta \frac{r}{r_{kk}} \tag{15}$$

By assumption 6, S_k/p_k is a constant (given by equation (3) in the case of lognormality); hence, α_k does not depend on p_k. We can write (13) more compactly as

$$P_k = \alpha_k p_k - \gamma_k a_k \tag{16}$$

Since γ_k is evidently positive, we have

$$\frac{dP_k}{da_k} = -\gamma_k < 0$$

More interestingly, γ_k (the absolute value of the slope of the demand function) is an increasing function of p_k, since equation (15) gives

$$\frac{d\gamma_k}{dp_k} = \frac{2\gamma_k}{p_k} \tag{17}$$

which is positive so long as p_k is positive.[5]

The model thus implies that an individual's demand function for a risky asset is linear and negatively sloped, and that the intercept and slope increase (in absolute value) with the asset's expected *EPV*.

C. *Implications about Market Behavior*

We introduce three additional assumptions about markets and individuals.

7. Securities markets are perfect in that 1) information is costless and transmitted instantaneously; 2) transactions are costless and can be made instantly; and 3) there are no institutional constraints upon transaction volumes or price changes.

8. An individual's interpretation of a bit of new information about a security normally depends in part on his current opinion about the asset. An individual who is bullish about a security, as a result of previous information, tends to disregard "small" bits of new information which another investor, already pessimistic about the security, regards as bearish. Similarly, the bear discounts minor information which the bull considers favorable.

9. At any moment the market for the kth asset consists of two numerically equal groups of investors, the "bulls" and the "bears," with the following characteristics: 1) each bull has a higher p_k than every bear; 2) bulls are no less risk averse than bears and no less bullish than the bears, on average, about the *other* $n-1$ risky assets; and 3) bulls and bears perceive the same correlations matrix of *EPV*s of the n risky assets and the same coefficients of variation.

The assumptions characterize the kth market as perfect and as composed of two equal groups of investors. The $J/2$ investors with the highest p_k are bulls; the rest are bears. Assumption 8 states that good news leads bulls to revise upward their expectations of the kth asset's *EPV* but that such news is ignored by bears. Bad news, while rejected by bulls, causes those who are already relatively pessimistic about p_k to lower their estimates even farther. This assumption is consistent with the observation that individuals tend to seek out, remember, and accept more readily information which reinforces their current views.

---

[5] In the case of lognormal $\tilde{p}_k$ it is easy to see also that $d\gamma_k/d\mu_k = 2\gamma_k$.

A given grouping of investors in any market is maintained so long as information is evaluated in this manner, since the flow of new information merely serves to make the bulls more bullish and the bears more bearish. We may suppose, however, that this reinforcing-interpretative behavior is abandoned temporarily under the influence of large news shocks. Shocks large enough to change opinions cause bulls and bears to break ranks and may produce entirely new groupings of investors.

From these assumptions and the theory developed in Sections IA and IB, we can now derive implications about market prices and transaction volumes. From equation (16) the demand function of the jth investor is

$$P = \alpha_j p_j - \gamma_j a_j$$

(For compactness we have dropped the subscript k, its being understood that this and subsequent expressions hold for any one of the n risky assets.) Suppose there is a total of J investors. Since the a_j may be negative (assumption 4), the market demand function is also linear. It is given by

$$(18) \qquad P = \left(\sum_{j=1}^{J} \frac{1}{\gamma_j}\right)^{-1} \sum_{j=1}^{J} \frac{\alpha_j p_j}{\gamma_j} - \left(\sum_{j=1}^{J} \frac{1}{\gamma_j}\right)^{-1} A$$

where $A \equiv \sum a_j$. Making the partial equilibrium assumption that (small) changes in the price of the asset have negligible effects on other prices, we can regard P as determined by equation (18) together with a constant supply function, $A = A_0$.

Separate demand functions can be derived for each of the two groups of investors specified by assumption 9. It will be convenient to write these functions in the following simple form:

$$(19) \qquad P = \Psi_T - \Phi_T a_T$$

$$(20) \qquad P = \Psi_U - \Phi_U a_U$$

where the subscripts T and U (from Latin) denote "bull" and "bear," respectively; where

$$(21a) \qquad \Phi_T \equiv \left(\sum_{j \in T} \frac{1}{\gamma_j}\right)^{-1} > 0$$

$$(21b) \qquad \Psi_T \equiv \Phi_T \sum_{j \in T} \frac{\alpha_j p_j}{\gamma_j}$$

and where Φ_U and Ψ_U are analogously defined. The terms Ψ_T and Ψ_U are evidently weighted averages of the intercepts, $\alpha_j p_j$, of the demand functions of individuals in the respective groups.

Equations (19) and (20) yield a new expression for market demand:

$$(22) \qquad P = \frac{\Psi_U \Phi_T + \Psi_T \Phi_U}{\Phi_T + \Phi_U} - \frac{\Phi_T \Phi_U}{\Phi_T + \Phi_U} A$$

where $A = a_T + a_U$.

From definitions (15) and (21) it can be seen that both Φ and Ψ increase under a proportional increase in p for all members of a group. From this it follows that if good news enters the market, causing an equiproportional increase in each of the p_j, $j \in T$, but not affecting the p_j, $j \in U$, then the intercept of the market demand function and the absolute value of its slope both increase. Given some stock supply function, $A = A_0$, we would expect the increase in the demand function's intercept and slope to cause, as in Figure 1, an increase in market price P.[6]

---

[6] Figure 1 also shows, however, that the increase in p would lead to a *decrease* in price if A_0 were sufficiently large. This apparent paradox can be understood by

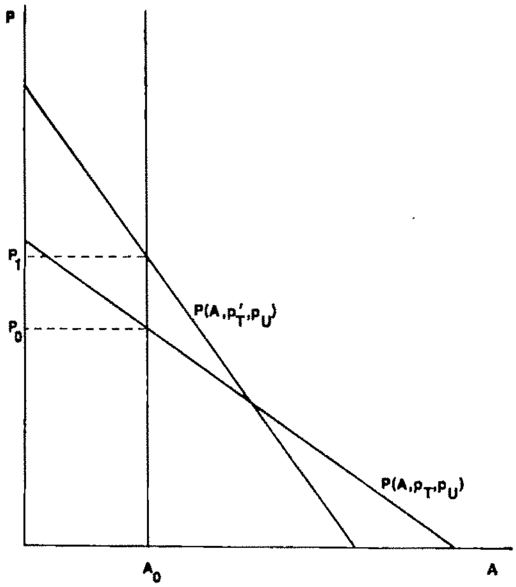


FIGURE 1. THE EFFECT OF GOOD NEWS ON MARKET DEMAND AND MARKET PRICE

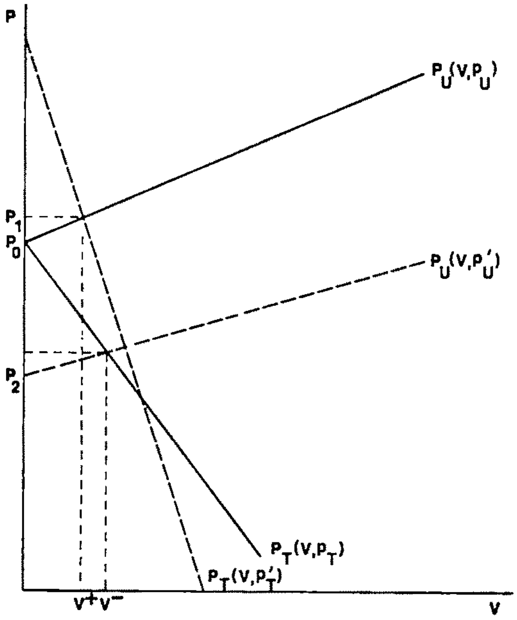


FIGURE 2. PRICE CHANGES AND TRANSACTION VOLUMES RESULTING FROM GOOD AND BAD NEWS

Equation (22) allows us to determine the effects on market price of expectational changes, but it provides no information about transaction volume. To determine the number of shares traded following a news shock, we must derive the excess demand and supply functions of the two groups of traders. The bull excess demand function and the bear excess supply function jointly determine the new equilibrium price and the volume.

The bull excess demand function and bear excess supply function from equations (19) and (20), respectively, are

$$(23) \qquad P = (\Psi_T - \Phi_T a_{T0}) - \Phi_T V$$

$$(24) \qquad P = (\Psi_U - \Phi_U a_{U0}) + \Phi_U V$$

where V represents the number of shares traded, and where a_{T0} and a_{U0} are the number of shares possessed by the two groups before trading begins. The prices corresponding to quantities in parentheses are equal to the "null" prices[7] of the two groups: the prices at which they would be satisfied with their current holdings (or, equivalently, at which excess demands and supplies would be zero). By definition, in market equilibrium each group holds the quantity at which its null price equals market price, so that

$$P_0 = \Psi_T - \Phi_T a_{T0} = \Psi_U - \Phi_U a_{U0}$$

The solid lines in Figure 2 depict the bull demand and bear supply functions in initial equilibrium. From such an equilibrium at market price P_0 we shall exam-

---

examining the demand function of a single investor (equation (13)). For large a an increase in $\hat{p}$ would so increase the asset's contribution to the risk of the individual's portfolio that he would be willing to continue to hold a units only if P were to decline. For a single investor the demand function can be shown to rotate about its midpoint when $\hat{p}$ changes. The demand curve of a group of investors can be shown also to rotate about its midpoint when $\hat{p}$ changes equiproportionately for all members of the group. Naturally, the model has descriptive value only if we regard the holdings of the groups as being small enough (in the observed range of market price) that expectational changes do not have perverse effects on quantity demanded.

[7] The term is Kenneth Boulding's.

ine alternately the effects of good and bad news on market price and volume of transactions.

Consider first a minor news shock which the bears disregard but which leads each of the bulls to revise upward his estimate of p by the factor $\delta_T > 1$. The new market price after the reestablishment of equilibrium is given by either of the following equations:

$$P_1 = (\delta_T \Psi_T - \delta_T^2 \Phi_T a_{T0}) - \delta_T^2 \Phi_T V^+ \tag{25}$$

$$P_1 = (\Psi_U - \Phi_U a_{U0}) + \Phi_U V^+ \tag{26}$$

where V^+ is the number of shares exchanged following the receipt of good news. From equation (26) the *change* in price, $\Delta P^+ \equiv P_1 - P_0$, is

$$\Delta P^+ = \Phi_U V^+ \tag{27}$$

Consequently, the ratio of transaction volume to the change in price, following the receipt of good news, is simply

$$\frac{V^+}{\Delta P^+} = \frac{1}{\Phi_U} \tag{28}$$

The volume V^+, determined from equations (25) and (26), is

$$V^+ = \frac{(\delta_T \Psi_T - \Psi_U) - (\delta_T^2 \Phi_T a_{T0} - \Phi_U a_{U0})}{\delta_T^2 \Phi_T + \Phi_U} \tag{29}$$

Next, consider the effect of bad news, starting again from the equilibrium at P_0. This time the shock is ignored by the bulls but causes each of the bears to reduce his expectation of EPV by the factor $\delta_U^{-1} < 1$. The new market price P_2 is given by either of

$$P_2 = (\delta_U^{-1} \Psi_U - \delta_U^{-2} \Phi_U a_{U0}) + \delta_U^{-2} \Phi_U V^- \tag{30}$$

or

$$P_2 = (\Psi_T - \Phi_T a_{T0}) - \Phi_T V^- \tag{31}$$

and the change in price, $\Delta P^- \equiv P_2 - P_0$, is

$$\Delta P^- = -\Phi_T V^- \tag{32}$$

The volume of transactions accompanying the downtick in price is

$$V^- = \frac{\delta_U(\delta_U \Psi_T - \Psi_U) - (\delta_U^2 \Phi_T a_{T0} - \Phi_U a_{U0})}{\delta_U^2 \Phi_T + \Phi_U} \tag{33}$$

while the ratio of volume to price change is

$$\frac{V^-}{\Delta P^-} = -\frac{1}{\Phi_T} \tag{34}$$

The behavior of the excess demand and supply functions following good and bad news is depicted in Figure 2.

Since $\Phi_T > \Phi_U$ by definitions (15) and (21) and assumption 9, equations (28) and (34) imply that the ratio of volume to price change on upticks exceeds the absolute value of the ratio of volume to price change on downticks. Stated differently, the number of shares exchanged on a transaction in which price rises exceeds the volume accompanying a price decline of the same magnitude. This implication evidently depends heavily upon the assumption that the coefficient of variation of anticipated EPV is constant. It is this assumption which implies that bulls consider assets to be riskier than bears do, and it follows from this that the bulls' demand function is steeper than that of the bears. We now turn to the testing of this hypothesis.

II. Empirical Findings

In Section IIA below we discuss the selection of a sample of securities for testing the model and explain the treatment of exceptional observations. The test itself is described in Section IIB.

A. *Selection of the Data*

The test of the model was conducted with individual transactions data, which are available in the form of volume and price for each transaction in all bonds and stocks on the two major exchanges.[8] I chose to restrict the population of securities to a set which conformed as closely as possible to the market assumptions on which the model relies. Three considerations governed the restrictions.

First, the model requires that most shocks be small enough that changes in bull and bear expectations of *EPV* do not occur simultaneously. Yet in markets where transaction costs are high, large shocks may be required to generate transactions. Harold Demsetz has argued that one component of transaction costs in securities markets, the bid-asked spread, is lower for actively traded securities than for those in which transactions occur infrequently.[9] Consequently, we should expect the proportion of transactions generated by large shocks to decline with the transaction rate. This suggests that actively traded securities should be used for our tests.

A second consideration in identifying a population of securities deals with broker commissions, the other major component of transaction costs. Fortunately, these costs are the same for sales and purchases.[10] However, distortions which would bias the test of our model are introduced by the commission schedule applying to stocks. On the leading exchanges, minimum commission rates per share increase monotonically with sale price. This causes the asked price of sellers (bears) to increase with volume at a faster rate than otherwise, and the bid price of buyers (bulls) to decline at a lesser rate than in the absence of commissions.[11] (See Figure 3.) Whether or not the distortions imposed by stock commissions are sufficient to reverse the hypothesized relation between the slopes of bull and bear demand functions is an empirical question that will be addressed in another paper. For the first test of the model, however, I avoided the problem by restricting the security population to actively traded corporate bonds. Bond commissions, with minor exceptions, are flat fees.[12] In addition, they usually constitute smaller percentages of sale prices than do commissions on stocks. Another important advantage of bond transactions data is that they are free of the influences of specialists, who, as William Baumol argues, can act (within limits) as price setters.[13]

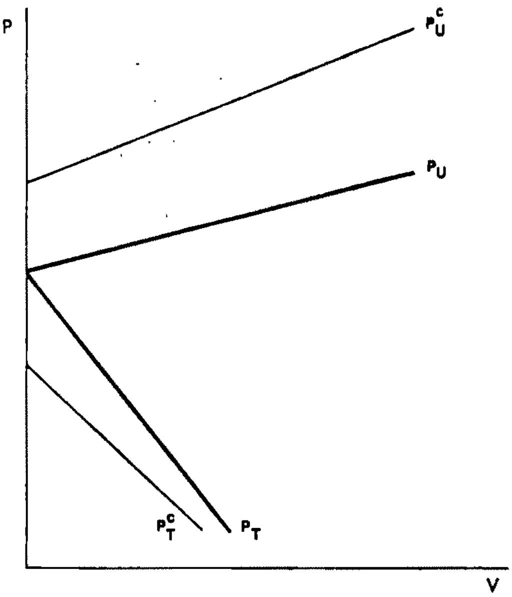


FIGURE 3. THE EFFECT OF STOCK BROKERAGE COMMISSIONS ON EXCESS DEMAND AND SUPPLY FUNCTIONS

---

[8] The data were obtained from Francis Emory Fitch.

[9] This finding is supported by the Report of the Special Study of Securities Markets, p. 75.

[10] The seller does pay, in addition to the commission, a small New York state transfer tax and an *SEC* fee which normally amount to less than .5 percent of the value of the transaction.

[11] This price effect has been offset to some degree by discounts on large stock transactions that have been in effect, in one form or another, since late 1968.

[12] The usual fee is $5 per $1000 par value. At least one major firm reduces its rate to $2.50, the *NYSE* minimum, for each bond after the first 49 traded in a single transaction. In addition, lower rates increasing by steps with price are authorized by the Exchange for bonds selling for less than $100. None of these was included in the population.

[13] On the New York Stock Exchange the most ac-

TABLE 1—SECURITIES INCLUDED IN THE SAMPLE

Name	Sample Size	Ticker Symbol
Chrysler Corporation 8 7/8s95	181	*C*
Chase Manhattan 4 7/8s93	87	*CMB*
Commonwealth Edison 8 3/4s75	131	*CWE*
Dow Chemical 8 7/8s2000	74	*DOW*
Ford Motor Co. 8 1/4s74	172	*F*
Ford Motor Credit 8 3/4s75	105	*FCR*
General Motors Acceptance Corporation 8 3/4s77	155	*GMA*
Standard Oil of New Jersey 6s75	167	*J*
J. C. Penney 8 7/8s95	61	*JCP*
National Cash Register 6cv95	259	*NCR*
New Jersey Bell Telephone 9.35s2010	156	*NJT*
New England Telephone and Telegraph 8 5/8s2009	188	*NTT*
Sears Roebuck 8 1/8s76	94	*S*
South Central Bell Telephone 8 1/4s2000	130	*SCB*
Santa Fe Industries 6 1/4cv98	108	*SFF*
Signal Companies 8.85s94	256	*SGN*
Standard Oil of Indiana 6s98	73	*SN*
Southwestern Bell Telephone 8 3/4s2007	171	*SWT*
American Telephone and Telegraph 8 3/4s2000	1010	*T*
Union Oil of California 8 1/4s76	97	*UCL*

The final consideration in the choice of a population is institutional limits on transaction volume and price change. Most bond price changes occur in multiples of $\frac{1}{8}$th percent of par, and transactions occur in multiples of $1000 par value. The only apparent way to limit the possible influence of these constraints is to choose high priced, widely held bond issues, for which the minimum price change and volume constitute small percentages of market price and number of bonds outstanding.

Based on these restrictions a population of 78 securities was identified, consisting of all *NYSE* bonds with "asked" price quotations exceeding 90 percent of par and for which at least 100,000 bonds were outstanding on the first day of the sample period. A sample of 20 bonds (Table 1) was selected randomly from the population and the complete sequences of prices and volumes recorded for all transactions in each bond during the month of January, 1971.

A number of exceptional transactions in the data required special consideration. To test the model, we must distinguish transaction volumes accompanying downticks from those accompanying upticks. However, for each bond in the sample there were many transactions in which price did not change at all from its previous level. Such cases could be treated in two ways: First, transactions in which price did not change could simply be dropped from the data. Then each remaining volume could be associated unambiguously with a price uptick or downtick. Alternatively, volumes accompanying all successive transactions at a given price could be added together. Then each of the *cumulative* volumes could be identified with a downtick or an uptick.

Both ways of treating these exceptional transactions appear to have some theoretical merit, as the following argument suggests. Several phenomena can account for the occurrence of two or more successive transactions at the same price level: 1) the existence of minimum price change constraints; 2) the queuing of orders at the same bid or offer price;[14] and 3) conflicting interpretations of news that cause bull excess demand and bear excess supply curves to shift in opposite directions by

---

tively traded bonds (called "free" bonds) are bought and sold by means of verbal bids and offers made by brokers in the "crowd." Transactions in so-called "cabinet" bonds, the less active issues, are made with written bids and offers. In neither case are specialists called on to make a market. See George Leffler and Loring Farwell, pp. 273–81. For a discussion of the specialist's freedom in determining the price at which a stock transaction is made, see the Special Study of Securities Markets, especially pp. 102 and 136.

[14] These may appear on the tape as separate transactions even if the broker executes them as a group.

offsetting amounts. The last factor by itself can scarcely be responsible for many of these transactions, since it depends entirely on coincidence; but the other factors may be important. If the first factor plays an important role in bringing about successive transactions at a single price, then volumes on zero-price-change transactions should be eliminated from the data, since the transactions may have been caused by different news shocks. But if the second factor is sometimes important, successive transactions at the same price may have resulted from a *single* news shock. In this case they should all be counted as one transaction, and the cumulative volume should be used in the tests.

Since it seemed impossible to choose between the two methods for handling zero-price-change transactions, the test was conducted with both noncumulative and cumulative volume data.

B. *A Test of the Model*

The model implies that $V^+/\Delta P^+ > V^-/|\Delta P^-|$, or equivalently that $\Delta P^+/V^+ < |\Delta P^-|/V^-$. A test which compares the means of these ratios over time is ambiguous, since in a sample containing extreme values it is possible for the mean of the $V^+/\Delta P^+$ to exceed the mean of $V^-/|\Delta P^-|$ *and* for the mean of $\Delta P^+/V^+$ to exceed the mean of $|\Delta P^-|/V^-$. For this reason I employed the nonparametric Wilcoxon test to compare the centers of the distributions of $V^+/\Delta P^+$ and $V^-/|\Delta P^-|$. (See Alexander Mood and Franklin Graybill, pp. 417–19.) In samples of the sizes employed here the test yields for each bond a standard normal "z" statistic, which measures the standardized difference in the centering of the two distributions.

The z statistics for the 20-bond sample are presented in Table 2. Negative values support the theory. Sixteen of them from the tests with noncumulative volume have the correct sign; seventeen from the tests with cumulative volume. Fisher's test[15] of the combined results indicates that the probabilities of obtaining these outcomes by chance are .031 and 2.2 x 10^{-4}, respectively.

TABLE 2—"z" STATISTICS FROM WILCOXON TESTS OF THE DIFFERENCE IN CENTERING BETWEEN THE DISTRIBUTION OF $V^+/\Delta P^+$ AND $V^-/|\Delta P|$

	Wilcoxon "z" Statistics[a]	
Bond[b]	Noncumulative Volume	Cumulative Volume
C	.550	− .037
CMB	− .427	− .591
CWE	1.852	− .281
DOW	.486	.434
F	− .647	− .890
FCR	− .632	.329
GMA	−1.917	−2.036
J	− .158	− .481
JCP	.553	.155
NCR	−1.885	−3.084
NJT	− .566	− .747
NTT	− .790	−2.066
S	− .766	− .667
SCB	− .277	− .116
SFF	−1.272	− .669
SGN	− .312	− .244
SN	− .383	− .809
SWT	− .184	−1.032
T	−2.513	−3.272
UCL	− .178	− .080

[a] The model implies negative z.
[b] For explanation of symbols, see Table 1.

III. Summary and Conclusions

We constructed in this paper a model of securities markets which predicts with some accuracy the behavior of bond price changes and transaction volumes. The model regards all transactions as occurring between two groups of investors, the "bulls" and the "bears." Assuming that subjective probable outcomes of end-of-period value have constant coefficient of variation and that interpretations of new information typically reinforce existing opinions, the model implies that the ratio of transaction volume to price change on upticks exceeds the absolute value of this ratio on downticks. This hypothesis was strongly supported by an empirical test with individual transactions data from a sample of widely held, actively traded, high priced corporate bonds.

---

[15] The following method was employed to construct a single significance test from the 20 independent samples: if the null hypothesis is true for each of n experiments, and if α_i is the probability value (significance level) for the ith experiment, then the quantity $-2 \sum ln\, \alpha_i$ has a Chi-square distribution with $2n$ degrees of freedom. See R. A. Fisher, pp. 99–101. W. A. Wallis provides additional information about the use and applicability of this test.

Several experiments were performed to assess the sensitivity of the model to institutional constraints on price change and volume and to the presence in the data of transactions generated by large shocks. While not conclusive, the results suggested that the constraint on volume may be important. For example, in one experiment the Wilcoxon tests were repeated with a second sample of 20 bonds from another population. For this sample the average number of securities outstanding was only about 30 percent of the average for the first sample. Although these data supported the tested proposition, the results were much less convincing than those reported in Table 2.[16] Since the inability to sell or buy *fractions* of one bond presumably constitutes a greater departure from the perfect-market assumption (assumption seven) the smaller the average holding, the fact that "thinner" issues did not support proposition two as strongly may mean it is sensitive to the constraint on volume.

The sensitivity of the model suggests in turn that the slopes of the demand functions of the bullish and bearish sectors of the market do not differ markedly. Under the model's assumptions the difference between these slopes would be small if and only if there were little diversity of opinion among investors. In the United States, where corporations are required to disclose certain financial information and where advisory services disseminate this information rapidly, one suspects that the average expectations of bulls and bears are not usually far apart. In this context, it would be interesting to repeat the test of this model with pre-1934 *U.S.* data and with transactions data from a nation with highly developed financial markets but no disclosure requirement. The first test especially would provide evidence bearing on George Benston's thesis that the *SEC* Act of 1934 has done little to reduce uncertainty in financial markets.

REFERENCES

J. Aitchison and J. Brown, *The Lognormal Distribution*, Cambridge 1966.

W. Baumol, *The Stock Market and Economic Efficiency*, New York 1965.

G. Benston, "Required Disclosure and the Stock Market: An Evaluation of the Securities Exchange Act of 1934," *Amer. Econ. Rev.*, Mar. 1973, *63*, 132–55.

G. Bierwag and M. Grove, "On Capital Asset Prices: Comment," *J. Finance*, Mar. 1965, *20*, 89–93.

K. Boulding, "A Liquidity Preference Theory of Market Prices," *Economica*, May 1944, *9*, 55–63.

H. Demsetz, "The Cost of Transacting," *Quart. J. Econ.*, Feb. 1968, *87*, 33–53.

T. Epps and M. Epps, "The Stochastic Dependence of Security Price Changes and Transaction Volumes: Implications for the Mixture-of-Distributions Hypothesis," *Econometrica*, forthcoming.

D. Farrar, *The Investment Decision under Uncertainty*, Chicago 1967.

R. Fisher, *Statistical Methods for Research Workers*, 14th ed., New York 1970.

---

[16] Fisher's test of the combined results yielded probability values of .131 for cumulative data and .271 for noncumulative data. Details of this and the other sensitivity tests will be furnished upon request.

F. E. Fitch, *Bond Sales on the New York Stock Exchange, A Daily Market Publication,* New York.

M. Godfrey, C. Granger, and O. Morgenstern, "The Random-Walk Hypothesis of Stock Market Behavior," *Kyklos,* 1964, *17,* 1–30.

C. Granger and O. Morgenstern, "Spectral Analysis of New York Stock Market Prices," *Kyklos,* 1963, *16,* 1–27.

G. Leffler and L. Farwell, *The Stock Market,* 3d ed., New York 1963.

J. Lintner, "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *Rev. Econ. Statist.,* Feb. 1965, *47,* 13–37.

A. Mood and F. Graybill, *Introduction to the Theory of Statistics,* 2d ed., New York 1963.

A. Moore, "Some Characteristics of Changes in Common Stock Prices," in P. Cootner, ed., *The Random Character of Stock Market Prices,* Boston 1964, 139–61.

J. Mossin, "Equilibrium in a Capital Asset Market," *Econometrica,* Oct. 1966, *34,* 768–83.

M. Osborne, "Brownian Motion in the Stock Market," *Operations Research,* Mar./Apr. 1959, *7,* 145–73.

———, "Periodic Structure in the Brownian Motion of Stock Prices," *Operations Research,* May/June 1962, *10,* 345–79.

W. Sharpe, "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," *J. Finance,* Sept. 1964, *19,* 425–42.

W. Wallis, "Compounding Probabilities from Independent Significance Tests," *Econometrica,* July/Oct. 1942, *10,* 229–48.

C. Ying, "Stock Market Prices and Volumes of Sales," *Econometrica,* July 1966, *34,* 676–85.

U.S. Congress, "Report of the Special Study of Security Markets," 88th Congress, House Docket no. 95, Part 2, Washington 1963.

The Measurement and Trend of Inequality: A Basic Revision

By MORTON PAGLIN*

The Lorenz curve has for some sixty years dominated our approach to the measurement of income inequality, and in consequence has become one of the most widely disseminated concepts in economics. This paper will attempt to show that a profound distortion of the degree of income inequality has resulted from the use of the Lorenz curve and the related Gini concentration ratio, both of which combine and hence confuse *intra*family variation of income over the life cycle with the more pertinent concept of *inter*family income variation which underlies our idea of inequality and economic class. To remedy the defects of the Lorenz-Gini approach, a fundamental revision is proposed which takes as its point of departure the replacement of the 45 degree line of equality with a new function generated on the basis of a more careful and explicit definition of perfect equality. This will allow us to utilize annual income data while avoiding the pitfalls now associated with its use. The widely accepted estimates of current inequality, and of historical changes in Gini ratios, are shown to have misrepresented an important characteristic of the *U.S.* economy.

I. Alternatives to Lorenzian Equality

The principal weakness of the Lorenz concept lies buried in the implications of the line of perfect equality. While a few writers have remarked that the 45 degree line has only mathematical significance, they have along with other users of the concept thrust upon it a considerable normative burden. There is really no way to avoid this if the Lorenzian area of inequality is to have any meaning as an index of income distribution. The basic fault with the existing line of equality is that it overspecifies the conditions of equality when used with annual income data. Assuming for the moment no economic growth, the maintenance of Lorenzian equality requires not only equal lifetime incomes, but additionally that families during the period of child rearing, when they have maximum income needs, must have the same incomes as families in the retirement stage of the life cycle when they have minimum economic responsibilities and maximum assets. These conditions are specified by the 45 degree line because families of all ages must have equal incomes in any given year, and this can be realized only if all families have perfectly flat age-income profiles. Yet it would be difficult to argue that a flat age-income profile is essential to equality, and there are substantial reasons for maintaining that such a mechanical requirement is itself highly inequitable. In Figure 1A, average *U.S.* family income by age of family head is indicated by curve *AB* and compared with the flat mean income profile *CD* specified by the Lorenzian equality line. It is certainly reasonable and sufficient to define perfect equality as a condition in which all families have equal lifetime incomes but traverse path AB rather than CD. In fact, to insist on flat age-income profiles flies in the face of both consumption needs and the realities of income production, for we know that investment in human resources through education and training produces a more peaked age-income profile, even apart from lifetime income differences (note Figure 2A). Lorenzian equality also does not allow family income to be increased by additional members entering the labor force. Few economists would explicitly postulate the above conditions as the requirements of perfect equality, yet this is the normative reference level embodied in the Lorenz definition. It is the standard by which we measure inequality and thereby subjectively judge the equity of our economic system.

---

* Professor, department of economics and Graduate Faculty of Urban Studies, Portland State University. Leonard Gionet and Peter Mann assisted with the statistical work and Peggy Davidson with the computer programming.

Some of the limitations of the Lorenz-Gini measure have been recognized in the literature, but the remedies suggested have taken us on a detour of data manipulation: the use of hypothetical lifetime incomes, age-group Gini coefficients, etc. However, none of these approaches has had much impact on our empirical estimates of the trend of inequality, for reasons to be discussed below. Apparently, no one has conceived of the alternative approach: reconstructing the reference line of equality to match the excellent annual income data at our disposal.

Once we cast aside the socially unrealistic 45 degree line of equality, we are free to generate new reference lines corresponding to explicit and reasonable definitions of equality, equity, or Pareto optimality. These new standards will be called P-reference lines, and in the balance of this paper one particular definition will be emphasized and used to make new estimates of inequality in the period 1947 to 1972. The new P-reference line is defined in a way which conforms to what casual users of the L curve might infer is the meaning of equality: equal lifetime incomes but not the added constraint of a

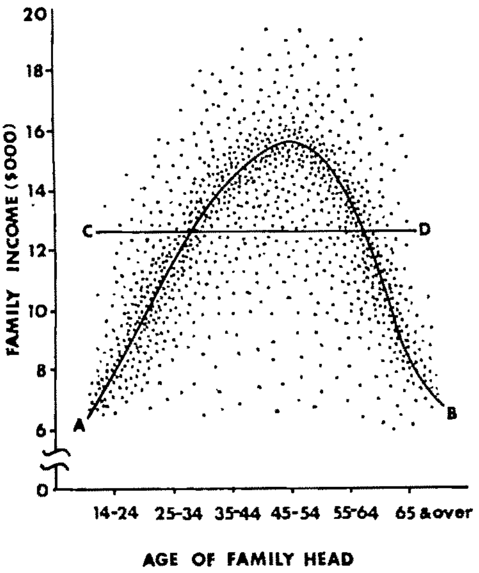


FIGURE 1A. 1972 AVERAGE AGE-INCOME PROFILE (Dots are only suggestive of distribution)

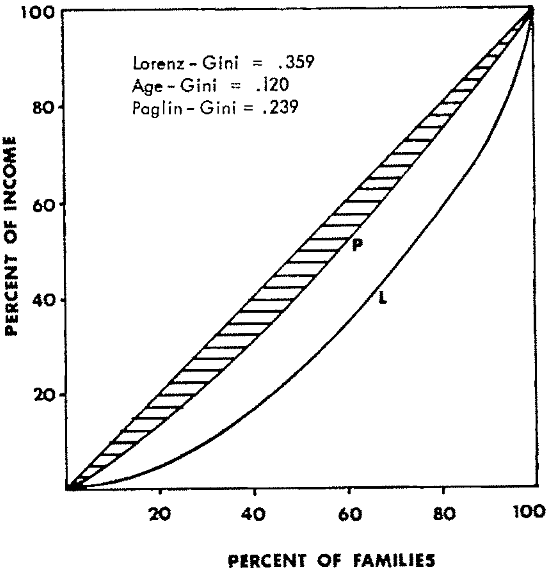


FIGURE 1B. 1972 CUMULATIVE INCOME DISTRIBUTION SHOWING AGE-RELATED INEQUALITY

Source for both figures: CPR, no. 90, pp. 51–53.

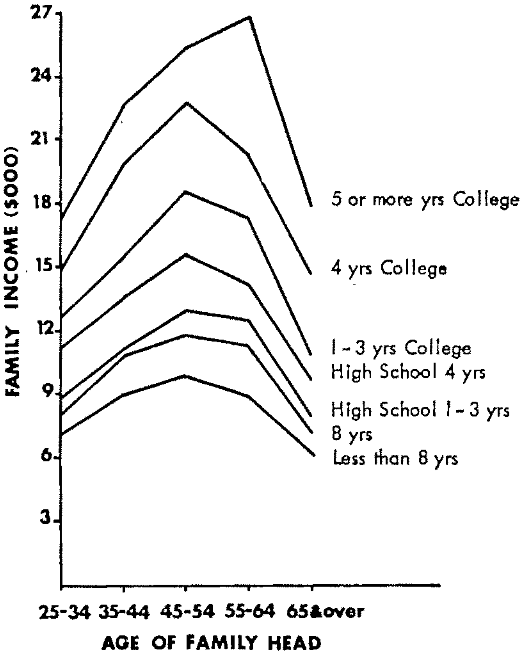


FIGURE 2A. AGE-EDUCATION PROFILES

Source: CPR, no. 90, pp. 79–81.

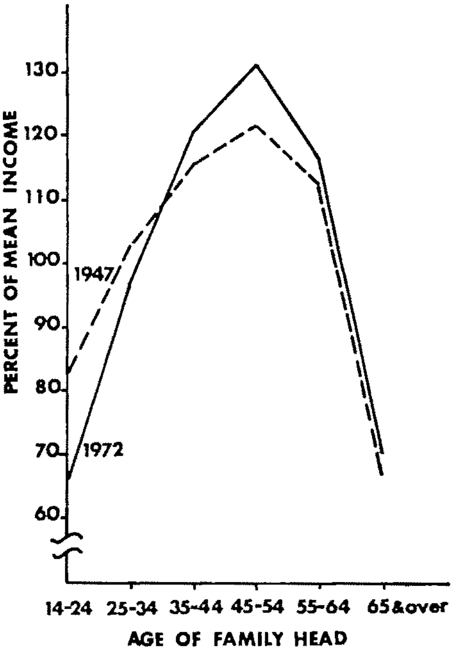


FIGURE 2B. INCOME PROFILE INDEXES

Source: Trends in the Income of Families and Persons in the United States: 1947–1964, p. 187; *CPR*, no. 90.

flat age-income profile.[1] While income would vary over the life cycle as indicated by curve *AB*, every family at a given stage of its life cycle would have exactly the same income as others at that stage—that is, all would have the same age-income profile. (We measure the stages of the life cycle by the age of the family head.) Now, taking the *U.S.* family population—its age distribution and average age-income profile—what would such a distribution of income look like on a Lorenz diagram? This is of course our new *P*-reference line of equality. We can derive it by taking average family income in each age group and ranking these groups by their mean incomes, remembering that every member of a group is assumed to have the same income. As expected, the young families and the over 65 age group rank at the bottom of the income scale. By using the percentage each group represents in the total population of families, a Lorenz-type curve was constructed from data in Table 1 and shown in Figure 1B as the *P*-reference line.[2] The hatched area between the 45 degree line and the *P* curve can be measured by a Gini ratio which is hereafter designated as the age-Gini. A standard Lorenz curve of family income (*L*) was then added in Figure 1B. The new area of inequality between the *P* curve and the *L* curve can also be measured by a Gini ratio which will be referred to as the Paglin-Gini, while the traditional Gini ratio based on the 45 degree line will be called the Lorenz-Gini. The Lorenz-Gini

---

[1] Equal lifetime incomes are specified in a no growth economy; with growth, the definition requires equal lifetime incomes only for members of the same age group or generation. This point is discussed below.

[2] A cubic spline function (cited in the Appendix) was fitted to the cumulative data points to derive the *P* curve.

TABLE 1

Age Group (1)	Mean Income (2)	Number of Families (thousands) (3)	Income of Group (millions) (4)	Cumulative Percent Families (5)	Cumulative Percent Income (6)
14–24	$ 7,892	4,194	$ 33,099	7.7	4.8
65 and Over	8,356	7,590	63,422	21.6	14.1
25–34	11,699	11,941	139,698	43.6	34.1
55–64	13,757	8,677	119,232	59.6	51.8
35–44	14,394	10,723	154,347	79.3	74.3
45–54	15,690	11,258	176,638	100.0	100.0
Totals		54,373	686,436		

Note: The *P*-reference curve is a Lorenz-type curve drawn up as illustrated above for 1972 family incomes (*CPR*, Series P-60, no. 90, pp. 51–53), with age groups ranked by mean incomes, not by ages.

minus the age-Gini equals the Paglin-Gini.[3]

Comparing the *P* and *L* curves in Figure 1B, and the Gini ratios of the areas shown, one is struck by the fact that fully one-third of the total area of inequality falls between the 45 degree line and the *P* curve. This shows what a large amount of ballast is included in the Lorenz-Gini measure, for clearly the age-income profile and the age structure of the population are not related to the long-run or lifetime degree of inequality in the economic system. In any case, the question of the optimum age-income profile is a different issue from that of equality as commonly conceived; our technique allows these two questions to be considered separately. Looking at the *P* curve, it is evident that rigidly egalitarian assumptions (i.e., equal lifetime incomes) may nevertheless produce a sizable area of inequality when annual income data are used. Hence, the Lorenz-Gini ratio (.359) is 50 percent greater than the Paglin-Gini (.239) which more closely approximates a measure of long-run interfamily inequality.

II. Traditional Attempts to Improve the Lorenz-Gini Measure

For some years it has been realized that the use of annual income data exaggerates the area of inequality. The suggested remedy has been the generation of hypothetical lifetime incomes since actual lifetime income data do not exist. But even if we had the data (historical and projected)

---

[3] In graphic terms, the Gini ratio is the Lorenzian area of inequality divided by the total area of the lower right triangle (Figure 1B). Algebraically, it is another measure of variance: $G = \Delta_1/2u$ where u is the mean income and

$$\Delta_1 = \frac{1}{N(N-1)} \sum_{j=1}^{n} \sum_{k=1}^{n} \{ | x_j - x_k | f_j f_k \}$$

with x representing the values of the variate and f the frequencies (see G. U. Yule and M. G. Kendall, p. 146). Δ_1 is Gini's coefficient of mean difference; it is the summation of the absolute differences between each value of the variate and every other divided by the number of such pairings. The Paglin-Gini is simply: $P.G. = (\Delta_1 - \Delta_1')/2u$ where Δ_1' equals the mean difference of the *P* curve distribution, namely, the hypothetical equality distribution in which all family incomes lie on curve *AB*. The Gini coefficient of mean difference and the concentration ratio are comparable to the more familiar measures of dispersion σ and $100\,\sigma/u$, and are monotonically related to them. Referring to Figure 1A, the Lorenz-Gini coefficient may be compared to the total variance (σ^2) around the mean, *CD*. The hatched area in Figure 1B (age-Gini) is conceptually comparable to the variance of *AB* from *CD*, while the Paglin-Gini coefficient measures the variation of values from *AB*. Education produces a greater arching of *AB* and thereby increases Lorenzian inequality which is measured from *CD*, but not the Paglin-Gini which is based on *AB*. The critique of the Lorenz-Gini measure developed here also applies to most other measures of income variation insofar as they use a single reference point such as the mean.

to construct lifetime incomes by consuming units, and converted these to constant dollars, the resulting Lorenz curve would show far more, rather than less, inequality. For in a dynamic economy with annual growth in real per capita income of 2 percent, there will be very large differences in lifetime incomes of older workers and young workers entering the labor force. With an average spread of 45 years between the new entrants and the retiring group, we can expect the former will have lifetime incomes well over 100 percent greater than the latter, and there is no practical redistribution scheme which would enable the older workers to approach the probable lifetime incomes of the younger. Also, given the upward trend in the labor-force participation rate of women, lifetime family incomes will increase even faster than the incomes of family heads. In a stationary economy, lifetime income equality is a workable hypothesis; in a growth economy, it becomes an almost unattainable goal. Therefore, to suggest that we retain the Lorenz 45 degree line but define it in terms of *lifetime* income equality gets us further from reality. While it eliminates the specification of a flat age-income profile, it saddles us with another unreasonable requirement: intergenerational lifetime income equality. This postulate, added to the empirical unreality of lifetime income data, reduces interest in this approach.

In contrast, the new P-reference function avoids these cumbersome problems. It recognizes that the past is irretrievably past, and that older generations cannot catch up with their younger contemporaries. It defines perfect equality at any point in time as equal incomes for all families at the same stage of their life cycle, but not necessarily equal incomes between different age groups. Annual income differences between age groups are solely a function of the average age-income profile; this reflects society's need for varying income over the life cycle as well as other basic facts relating to productivity, investment in human resources, and the work-leisure preferences of households, but only in an average way insofar as these factors express themselves through the age variable. The maintenance of P-reference equality posits that all members of a generation (or narrow age group) would have the same lifetime incomes; differences in lifetime incomes between generations would be mainly attributable to secular growth in real income, and such differences would be compatible with P-reference equality.

A second approach to the deficiencies of the Lorenz-Gini coefficient involves the use of age-specific Gini. The empirical coefficients available are not really specific by age of family head but in fact represent broad age groups. This introduces spurious income variance by not fully eliminating the effect of the age-income profile. However, even if we had truly age-specific Gini, we would have the problem of weighting and combining fifty-some measures into one coefficient. The use of the P curve is simpler, and allows us to graphically separate intrafamily and interfamily inequality; it also provides us with a new standard for evaluating quintile shares.

III. The Question of Consuming Units: What Should be Made Equal?

In organizing data for a Lorenz curve, the question of defining the appropriate consuming unit is usually considered a minor technical problem and hence glossed over by economists who are not specialists in this area.[4] However, the question, "What units should be made equal under perfect equality?" has substantial effects on the area of inequality, second only to redefining the equality line itself. In Table 2, the Lorenz-Gini and Paglin-Gini concentration ratios are shown for three alternative consuming units. All are based on 1972 census data, but the Lorenz-Gini coefficients vary by as much as 29 percent, for example, between "households" and "persons in families." At one extreme, the use of households (which include single person units) assumes that equality is realized if all households (no matter what their size) have the same income; at the other extreme, equality is defined in terms of persons in families, which assumes that children are equivalent to adults and that there are no economies of scale in family units. When we use persons in families instead of number of family units, the inequality coefficient decreases because family size is positively correlated with income, not inversely as commonly thought.[5] The lower income families tend to be clustered more heavily at both ends of the age-income profile where family size is below average.

---

[4] One need only refer to our leading introductory economics text (Paul Samuelson, p. 85) which after nine editions muddies the issue by using a Lorenz diagram labelled "per cent of people" to represent tabular data showing "per cent families" as the consuming unit, thus overstating the area of inequality by 13 percent.

While the Lorenz-Gini measures vary by 29 percent, the difference increases to 92 percent when we include the Paglin-Gini ratios in the comparison. Two factors account for this tremendous range (.208 to .400): 1) the alternative definitions of the equality line in terms of a flat or curved age-income profile, and 2) the question of the appropriate consuming unit—households, families, or persons? Arguments have already been given for rejecting the gross Lorenz-Gini ratios in the first column of Table 2. Of the Paglin-Gini ratios in the third column, the household figure is the least acceptable since it lumps single person units along with the larger consuming units. This leaves us with the values for "families" (.239) and "persons in families" (.208) as defining a reasonable range for net interfamily inequality in the United States. These figures are a far cry from the usually published estimates of .35 to .40 and illustrate how easily the equality issue can be overblown. A very substantial part of the traditional area of inequality (one-third to one-half) is simply a function of the diversity in the ages and size of families, and the lifetime income pattern typical of a technically advanced society. Such inequality does not represent a limitation on lifetime opportunities, nor is it a quintessential evil to be obliterated if our society is to be considered just and humane.

TABLE 2—U.S. 1972 INEQUALITY COEFFICIENTS

Consuming Unit	Lorenz-Gini	Age-Gini	Paglin-Gini
Households: includes single person units and incomes of nonfamily members living in unit.	.400	.151	.249
Families: two or more persons (excludes unrelated persons in living unit).	.359	.120	.239
Persons in families.	.320	.112	.208

Sources: For Households, *CPR*, no. 89; for Families, *CPR*, no. 90; for Persons in families, *CPR*, no. 90, Table 2, and *Family Composition*, p. 55 for mean size of families, by age of family head. (All income data used in this article are before taxes, and represent money incomes only.)

IV. A Reassessment of the Trend of Inequality: 1947–1972

One major error resulting from the use of the Lorenz-Gini ratio has been the widely accepted conclusion that there has been no significant reduction of inequality from 1947 to 1972 despite the massive spending on education and training programs, the more generous cash and merit good transfers, and the legislative and judicial actions directed at bringing minor-

---

[5] *Current Population Reports*, P-60, no. 90, Table 1, p. 27. Family size increases over most of the income range, namely from $3,000 to $50,000.

TABLE 3—GINI CONCENTRATION RATIOS AND SHARE OF LOWEST QUINTILE: 1947–72

Year (1)	Lorenz-Gini (2)	Age-Gini (3)	Paglin-Gini (4)	Actual Share of Lowest Fifth (5)	Equality Share Using 20 Percent Point of *P* Curve (6)	Actual Share Divided by *P*-Equality Share (7)
1947	.378	.075	.303	5.1	15.5	.329
1948	.369	.076	.293	5.0	14.8	.338
1949	.379	.076	.303	4.5	15.1	.298
1950	.375	.077	.298	4.5	14.9	.302
1951	.361	.072	.289	4.9	14.8	.331
1952	.374	.078	.296	4.9	14.8	.331
1953	.360	*n.a.*	*n.a.*	4.7	*n.a.*	*n.a.*
1954	.373	.084	.289	4.5	14.6	.308
1955	.366	.088	.278	4.8	14.5	.331
1956	.355	.083	.272	4.9	14.3	.343
1957	.351	.089	.262	5.0	13.7	.365
1958	.354	.093	.261	5.1	13.5	.378
1959	.366	.090	.276	5.1	13.6	.375
1960	.369	.091	.278	4.9	13.8	.355
1961	.376	.090	.286	4.8	14.4	.333
1962	.365	.100	.265	5.1	13.5	.378
1963	.360	.098	.262	5.1	13.6	.375
1964	.352	.096	.256	5.2	13.7	.380
1965	.357	*n.a.*	*n.a.*	5.3	*n.a.*	*n.a.*
1966	.348	*n.a.*	*n.a.*	5.5	*n.a.*	*n.a.*
1967	.355	.110	.245	5.4	13.0	.415
1968	.343	.110	.233	5.7	12.9	.442
1969	.348	.115	.233	5.6	12.8	.438
1970	.355	.114	.241	5.5	12.9	.426
1971	.356	.119	.237	5.5	12.7	.433
1972	.359	.120	.239	5.4	12.9	.419

Sources: Column (2), years 1947–64, *Trends in the Income of Families and Persons in the United States: 1947–1964*, pp. 182–87; Gini ratios for years 1965–72 were computed by the cubic-spline method from data on family income in *CPR*, Series P-60. (See Appendix for a description of the method used.) Column (3) lists the Gini ratios of the area between the 45 degree line and the *P*-reference curve computed each year from age-income data in same sources as Column (2). Column (4) equals (2) minus (3) and represents net interfamily inequality. Columns (5), (6), and (7) used the same data sources and *P* curves as columns (2) and (3).

Note: n.a. = Not available because median rather than mean incomes by age groups were given for those years.

ities and underprivileged groups into the mainstream of the economy. A fresh analysis of the income data using the *P*-reference curve indicates a considerable reduction of net inequality and a marked improvement in the share of the lowest quintile. Table 3 shows the historical data on inequality coefficients. Column (2) lists the standard Lorenz-Gini concentration ratios with only a slight decline in inequality evident. This forms the basis of the current view that inequality has not changed in the last twenty-five years. Column (3) lists the age-Gini ratios; these must be computed for each year since the age-income profile and the age structure change slowly over time. We may note a significant increase in the age-Gini from 1947 to 1972; this is related to the expansion of higher education which results in a greater arching of the average age-income profile (Figure 2), and to the increase in the percent of the aged and young adults in the population.[6]

---

[6] In 1947 young and old families (head below 25 years and over 65) together equalled 13.9 percent of all families; in 1972 they were 21.7 percent of the total.

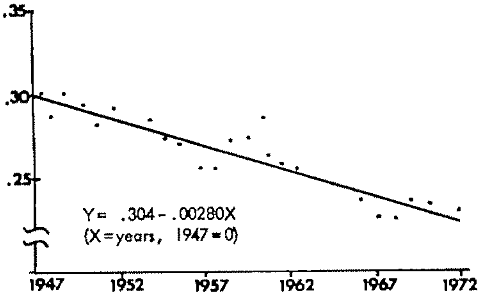


FIGURE 3A. TREND OF PAGLIN-GINI RATIOS

Note: $r=.94$; t=12.

Source for both figures: Table 3.

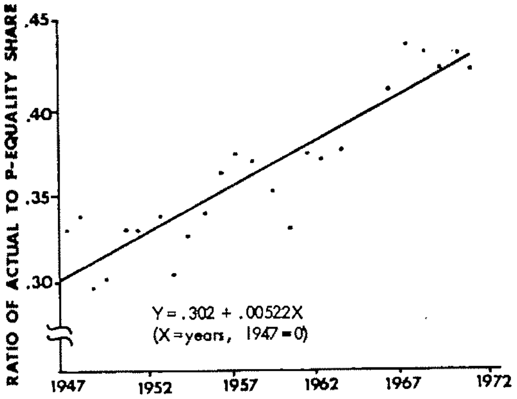


FIGURE 3B. TREND IN INCOME SHARE OF LOWEST QUINTILE USING *P*-REFERENCE CURVE

Note: $r=.91$; t=10.

Column (4) is the net measure of inequality expressed by the Paglin-Gini ratio, and is equal to column (2) minus (3). It reveals the decline in interfamily inequality of income, unobscured by changes in the age-income profile and in the age composition of the population. The least squares trend line fitted to column (4) data is shown in Figure 3A. In contrast to the traditional view, the equation indicates that inequality has declined 23 percent in the 25-year period, 1947–72.

V. Reevaluation of the Lowest Quintile

Let us consider the effect of the new equality reference line on the relative position of the lowest fifth of the income recipients. Using the traditional Lorenz curve (Table 4), the poorest families (in 1972) have 5.4 percent of the income pie whereas in a perfectly egalitarian society they would appear to get 20 percent of the income; hence it is implied that a measure of their deprivation is shown by the fact that they receive only 5.4/20 or 27 percent of the income share which would accrue to them under conditions of perfect equality. This grossly exaggerates the income deprivation of the lowest fifth. Using the more realistic definition of perfect equality embodied in our *P* curve, the relative position of the lowest quintile changes significantly, for we see that even under conditions of perfect equality of lifetime incomes (given the current age distribution and age-income profile), there would still be a "poorest" 20 percent of the families having only 12.9 percent of the income pie. Now it is appropriate to compare the 5.4 percent with the 12.9 percent, rather than with the 20 percent figure, unless we are prepared to argue that equality of lifetime income is not a reasonable norm, and that flat annual income from teenage to retirement and death is also somehow a necessary condition of equality. By employing the *P*-reference curve, we see that the poorest families receive 5.4/12.9 or 41.9 percent of the income which would accrue to them in an egalitarian utopia (Table 4). Hence, a revision of the equality norm has resulted in a 55 percent improvement in the relative assessment of the lowest income group, namely from 27 to 42 percent of the equality income share. (A similar analysis using the *P* curve as the standard results in a reduction of the share of the top

TABLE 4—1972 FAMILY INCOMES: LORENZ AND *P* CURVE EQUALITY COMPARISONS

Quintiles	Percent of 1972 Income (Actual)	Perfect Equality Lorenz Shares	Perfect Equality Shares Using *P* Curve	Actual Income Share as a Percent of: Lorenz Share	Actual Income Share as a Percent of: *P*-Curve Share
Lowest	5.4	20	12.9	27.0	41.9
2	11.9	20	17.8	59.5	66.9
3	17.5	20	21.5	87.5	81.4
4	23.9	20	22.9	119.5	104.4
Highest	41.3	20	24.9	206.5	165.9

Source: CPR, Series P-60, no. 90. *P*-equality shares interpolated from cubic-spline curve fitted to *P*-curve data points.

quintile from 2.07 times to 1.66 times the equality share, see Table 4.)

We may now begin to understand why the poorest 20 percent historically seem to be so resistant to improvements in terms of the conventional Lorenz curve analysis: to a large degree, the low percentage of the income pie going to this group is a built-in result of the age-income profile coupled with the age distribution of the population, and is not purely related to the condition of a permanent class of people excluded from the average level of real income enjoyed by most families. This has profound implications, for the failure to understand the nature and extent of economic inequality has led to a distorted assessment of our economic system.

This distortion has been further amplified by a statistical decision to exclude need-based in-kind transfers from the definition of income. As a matter of social policy, we have decided to mitigate poverty by making large transfers in the form of public housing, rent supplements, food stamps and food assistance, medicaid, and social services such as day care, etc. We then blithely exclude these transfers from the statistics on poverty and inequality and wonder about the lack of improvement in the share allotted to the lowest quintile! Fifteen years ago, in-kind transfers were small and could be overlooked, but now they exceed the cash allotments and are growing at a much faster rate. In fiscal year 1973, expenditures on the major income-tested transfer programs (federal, state, and local) totaled $30.7 billion of which 55 percent or $16.9 billion were in-kind transfers. While the cashing out of in-kind transfers is a complex task which will be dealt with in a subsequent article, a loose estimate of their distributive impact can be made. Since we are focusing on family units, only 80 percent of the in-kind transfers or $13.5 billion is applicable. Now assume that 20 percent of the $13.5 billion goes for administration. This leaves $10.8 billion as an addition to the income of the bottom quintile, boosting its 1972 income share from 5.4 to 7.0 percent. Hence if in-kind transfers are included as income, the lowest quintile has about 54 percent of the *P* curve equality share of 12.9 percent.[7]

Aside from the sharp increase of in-kind transfers which characterized the period 1947–72, the historical position of the lowest quintile purely in terms of money income has improved significantly. While this improvement has been obscured by the crude Lorenz approach (Table 3, column (5)), it is revealed by comparing in each year the actual share received by the lowest quintile with the *P* curve equality share. These ratios are shown in Table 3, column (7), and are plotted in Figure 3B along with the least squares trend. We may conclude that the overall reduction of inequality in the period 1947–72 has been accompanied by an upward trend in the money income share of the lowest quintile, namely, from 33 to 42 percent of the equality share.

---

[7] The data on transfer programs are derived from *The Budget of the U.S. Government, FY 1975, Appendix,* pp. 203–05, 440–42, 494, and 808. Programs such as medicare, which affect many low income families, have been excluded from the totals because they are not income tested. The 80 percent figure represents the proportion of low income persons living in families (*CPR,* no. 98, p. 13). The 20 percent for administrative costs is a rough estimate which may be low for existing programs but would be too high if a cash transfer strategy (negative income tax) were used.

VI. Wealth Inequality

From income inequality let us move on to the distribution of wealth. Our criticisms with some modification also apply to the Lorenz-Gini wealth coefficient. Simply stated, the problem is this: the 45 degree line of equality requires that wealth holdings among families be equal regardless of the age of the head of household. Given a typical age distribution of the population, this means that cumulative annual saving, plus interest compounded, can have no effect on individual family wealth. Again, this is an added specification which most persons would consider an unnecessary and unrealistic constraint even in an egalitarian society. It would seem sufficient to define equality as a social condition in

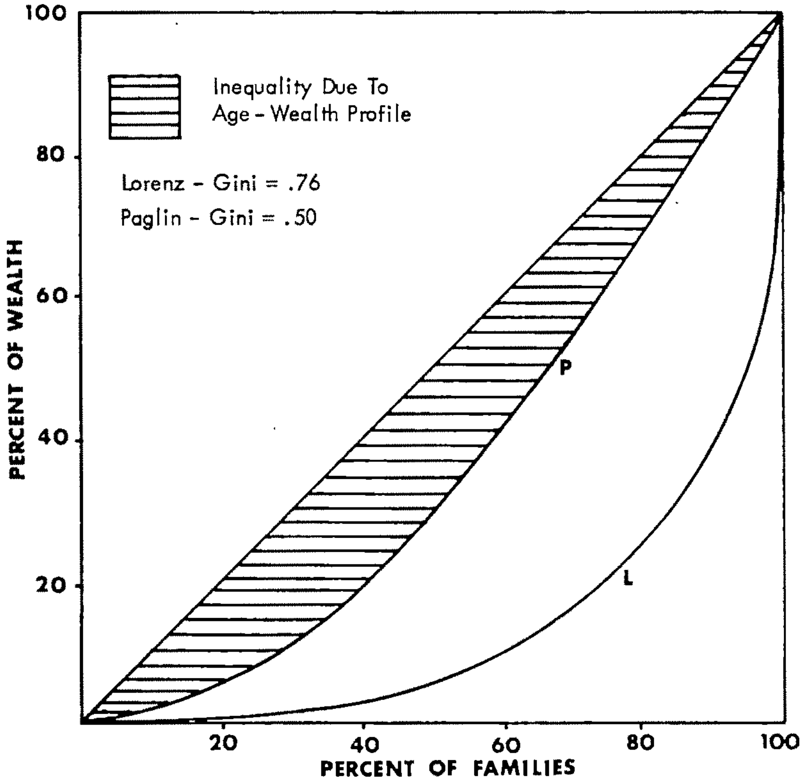


FIGURE 4. *U.S.* DISTRIBUTION OF WEALTH, 1962

Source: Projector and Weiss, pp. 110, 151.

which inheritance of wealth is not a significant factor in the *relative* distribution of wealth, all lifetime incomes (for a given generation) are equal, and family wealth is a function mainly of saving and time, with everyone having the same rate of saving (or dissaving) at a given stage of the life cycle. These assumptions would produce equality of wealth for families in the same age bracket, but would allow differences of wealth based on age. With this definition, what would a P-reference function for the United States look like, using the average age-wealth profile derived from Federal Reserve Board data (see Dorothy Projector and Gertrude Weiss)? The results are shown in Figure 4. The area between the 45 degree line and the P-reference function is hatched, and the Lorenz curve of wealth is also shown. It is apparent that a significant percentage of the total area of inequality (about one-third of it) would exist even in a rigidly egalitarian society, and therefore, this segment of wealth inequality reflects the age structure and the social savings function rather than fundamental (lifetime) inequality in the economic system. For the United States (in 1962) the Lorenz-Gini of wealth was .76 compared with a Paglin-Gini of .50; hence, the traditional measure has overstated the degree of interfamily inequality of wealth by about 52 percent.

VII. Summary and Conclusion

For over sixty years we have employed a measure of inequality—the Lorenz curve—based on an unrealistic standard of equality. When used with annual income data, Lorenzian equality requires a flat age-income profile as well as equality of lifetime incomes. While the use of lifetime income data would negate the overspecification of equality built into the 45 degree line, empirical and theoretical considerations make this approach impractical. By developing a new equality reference function, suitable for use with annual income data, we not only avoid cumbersome data problems, but we open up to explicit examination the meaning of perfect equality. We also separate the issue of *intra*family variation in income over the life cycle (accounting for one-third of Lorenzian inequality) from the more basic issue of *inter*family differences in lifetime incomes. An application of the new concepts to *U.S.* income and wealth data reveals that estimates of inequality have been overstated by 50 percent, and the trend of inequality from 1947 to 1972 has declined by 23 percent. On equity as well as efficiency grounds there is a substantial case for rejecting the Lorenzian standard of absolute equality. The overstatement of inequality has lent false urgency to the demand for rectification of our income distribution.

APPENDIX

The traditional method of calculating the Gini concentration ratio involves a straight line approximation between plotted points of the Lorenz curve, and a simple formula for determining the resulting area (see James Morgan, Appendix). This, however, underestimates the area of inequality, especially if the number of data points drops below eight, a common occurrence since we frequently have income shares by quintiles. There seemed to be two obvious remedies but both were found wanting: fit an appropriate curve type by least squares; this however didn't pass through all the data points and was therefore rejected. The second approach involved fitting an $n-1$ degree polynomial which passed through every data point but showed unpredictable flips between the points. What is wanted is a smooth, continuous function which will pass through all the points, whether few or many, and be usable for interpolation between the points. A cubic-spline function described in Brice Carnahan et al., p. 63, fulfilled these conditions, and was used to derive L curves and P-reference curves for all the years required. A computer program was written to generate about 40 interpolated points between every two actual data points, and by integration provide the area of inequality and the Gini ratio.

As a test of the program, we had the benchmark Gini ratios for 1947–64 (Census, *Trends in the Income of Families . . .*) computed from disaggregated data. Using quintile shares plus the share of the top 5 percent for these 18 years, Gini ratios were computed by both the straight line and the cubic-spline methods, and compared with the Census coefficients. The straight line method produced underestimations averaging 3.5 percent. The cubic-spline method yielded correct estimates (to three significant figures) for one-third of the years, and for the entire sample produced an average error of only 0.6 percent (.002 Gini point). The smooth curves and the interpolations provided by the program also made possible meaningful comparisons of various points on the P and the L curves.

REFERENCES

B. Carnahan, H. A. Luther, and J. O. Wilkes, *Applied Numerical Methods,* New York 1968.

J. Morgan, "The Anatomy of Income Distribution," *Rev. Econ. Statist.,* Aug. 1962, *44,* 270–83.

D. Projector and G. Weiss, *Financial Characteristics of Consumers,* Washington 1966.

P. A. Samuelson, *Economics,* 9th ed., New York 1973.

G. U. Yule and M. G. Kendall, *Introduction to the Theory of Statistics,* 14th ed., London 1957.

U.S. Bureau of the Census, *Current Population Reports,* (*CPR*) Series P-60, Nos. 51, 53, 59, 66, 75, 80, 85, 89, 90, and 98, Washington 1965–75.

———, *Trends in the Income of Families and Persons in the United States: 1947–1964,* tech. pap. no. 17, Washington 1967.

———, *Census of Population, 1970,* Vol. PC(2)–4A *Family Composition,* Washington 1973.

U.S. Office of Management and Budget, *The Budget of the U.S. Government, FY 1975, Appendix,* Washington 1974.

A Keynes-Friedman Money Demand Function

By Paul A. Meyer and John A. Neri*

Our main objective is to shed light on a question Milton Friedman regards as central: "Are transactions variables or asset variables most important in determining the demand for money?" (1970, p. 234). The recent theoretical literature on money emphasizes that money exists because it reduces transactions and information costs and saves time.[1] Recent empirical estimates of money demand have explored the influence of the real wage and unemployment rate, two variables which reflect the value of money in terms of reduced transactions time. While Dutton and Gramm's introduction of transactions variables like real wages is welcome, a more direct empirical answer to Friedman's question would give priority to the measurement of transactions and asset "constraints" rather than opportunity cost variables such as real wages.

In Section I we hypothesize that the demand for real balances depends on short-run expected income or transactions, which is not well measured by current income. The model distinguishes between short-run and long-run expected income and associates the former with transactions/finance variables and the latter with asset variables. Construction of the expectations variables incorporates elements from the *General Theory*, Keynes' (1937) replies to his critics, and Friedman's (1957, 1959) consumption and money demand studies. Reliance on doctrinal history alone leaves some ambiguities and our historical perspective may be questioned. Moreover, since John Muth has shown that permanent income may be the optimal predictor of income in any future period, we derive in Section II the expectations variables from the analysis of autoregressive processes. Using a stochastic model somewhat more general than Muth's, we show that Friedman's permanent income approximates the optimal forecast of long-run but not short-run income. The cumbersome mathematics necessary to derive optimal predictors shows that the optimal forecast for next period, a modified exponentially weighted average of past income, corresponds with our short-run expected income construct based on doctrinal history. In the regressions of Section III the variable discriminating between long-run expected income (asset motive) and short-run income (transactions motive) is highly significant in the M^1 (currency plus demand deposits) equations, supporting the transactions role of money. The model compares favorably with other estimates. Our conclusions are briefly summarized in Section IV.

I. Formulation of the Keynes-Friedman Money Demand Function

The generally accepted Keynesian money demand function relates money balances to current income which measures transactions requirements, and to an interest

---

* Associate professor of economics, University of Maryland, and economist, Federal Energy Administration, respectively. We gratefully acknowledge the comments of Martin Bailey, Dilip Madan, an unknown referee, and the editor without associating them with any errors which may remain. The Computer Center of the University of Maryland provided computer time.

[1] See Dean Dutton and William Gramm and the references they cite.

rate mainly because of the speculative motive. We feel Keynes would not object to, and indeed would even welcome, our modification of the transactions demand. Money demand is a function of the current expectation of income in the next period. Assuming constant elasticities so the relationship is linear in the logarithm,

$$M_t^* = \alpha_0 + \alpha_1 y_t^e + \alpha_2 r_t \tag{1}$$

where M^* is desired real balances, y_t^e is short-run expected real income, and r is an interest rate. To repeat, although not written explicitly the variables are *logs*. We temporarily abstract from an error term.

Some minor defense of (1) may be necessary. Assume there is approximately a one-to-one relationship between total transactions and income measures.[2] Considering the transactions motive alone, it is intuitively no less reasonable to relate desired real balances to planned transactions rather than current transactions. Equation (1) is entirely in accord with Keynes' thought if, as G. L. S. Shackle noted perceptively, we "... look in a part of the canon which few economists seem able to endure the sight of—or else they have never heard of it. It is his reply to his critics" (p. 516). In the debate with Ohlin, Robertson, and others over liquidity preference versus loanable funds, Keynes introduced a fourth motive for holding money—the finance motive. He states, "I should not have previously overlooked this point, since it is the copingstone of the liquidity theory of the rate of interest" (1937, p. 667). Stated succinctly, the finance motive refers to balances held for planned activity or what Keynes calls current *ex ante* output, in contrast to transactions balances held for actual activity or current *ex post* output. Of the two, Keynes states that the former may be the more important.

Speculating briefly on the reasons for the omission of the finance motive in the *General Theory*, Keynes was both overly pessimistic and optimistic about economists' ability to ascertain how income expectations are formed. Keynes distinguished between short-term and long-term expectations of output. With regard to the latter, Keynes position was "we simply do not know" (1936, pp. 147–64). On the other hand, short-run expected output or current *ex ante* output is well approximated by recent output, so well approximated that in the *General Theory* Keynes thought it was not necessary to distinguish between short-run expected output and actual current *ex post* output (1936, pp. 50–51). Thus Keynes' views on expected output differ markedly from his views on expected interest rates. Individuals do not uniformly expect interest rates in the near future to equal current rates, and the normal or permanent or long-run expected rate is bounded by past experience.

To express (1) in terms of observable magnitudes, we assume that individuals form income expectations in a manner analogous to the generally accepted view of Keynesian interest rate expectations, as interpreted by Franco Modigliani and Richard Sutch, for example. First, the Keynesian individual estimates normal or long-run expected income. We restrict Keynes somewhat by assuming normal income is virtually identical to the Friedman-Cagan formulation. That is, assume that the logarithm of normal real income y_t^n is an exponentially weighted arithmetic average of the logarithms of past income y.

$$\begin{aligned} y_t^n &= \sum_{i=0}^{\infty} (1-\beta)\beta^i y_{t-i} \qquad 0 < \beta < 1 \\ &= (1-\beta)y_t + \beta y_{t-1}^n \end{aligned} \tag{2}$$

---

[2] While bank debits may be more appropriate than income measures in the transactions approach, we show in a paper underway that debits and income are highly correlated. Moreover, debits have virtually no welfare implications, and construction of a price deflator for debits would be extremely difficult.

Equation (2) implies that the level of normal income is a geometric weighted average of past incomes, which is similar to Friedman's arithmetic average when the weights are the same. Indeed, Friedman and Anna Schwartz (1936b, p. 57) have suggested (2) as a way of measuring permanent income, and their suggestion was followed by Edgar Feige.[3] Corresponding to the treatment of interest rate expectations, deviations of current income from the normal are expected to be temporary. Expectations are regressive; income in the next period is expected to return towards the normal. Algebraically, next period or short-run expected income is given by

$$(3) \qquad y_t^e - y_t = \lambda(y_t^n - y_t)$$

where $0 \leq \lambda \leq 1$. The expected change in income depends simply on the deviation of current income from the normal. Thus, the Keynes-Friedman model states that past incomes influence current expectations only through their effect on normal income, which depends on the parameter β but not λ.

Substituting normal income in the expected income equation yields[4]

$$(4) \quad y_t^e = (1 - \lambda\beta)y_t + \lambda\beta(1 - \beta) \sum_{i=0}^{\infty} \beta^i y_{t-1-i}$$

With the aid of the recursion formula in (2), one can interpret expected income as (a) a weighted average of current and lagged normal income, where the weights are $(1-\lambda\beta)$ and $\lambda\beta$, respectively; (b) a weighted average of current and normal incomes at period t, where the weights are $(1-\lambda)$ and λ; or (c) current income plus a proportion λ of Friedman's temporary income. However interpreted, expected income is a modified exponentially weighted average of past income, the weights declining exponentially one period lagged. For any β, current (lagged) income has a greater (smaller) weight in expected than normal income if $\lambda < 1$. When $\lambda = 1$, full adjustment back to normal is expected to occur next period.[5]

Returning to the original money demand function, we make two alternative assumptions about the relationship between desired and actual balances.

$$(5a) \qquad M_t^* = M_t$$

$$(5b) \qquad M_t - M_{t-1} = \gamma(M_t^* - M_{t-1})$$

The first equation in (5) makes the restrictive assumption that actual and desired balances are equal, while the second equation states that during a period t only a fraction, $0 < \gamma < 1$, of the percentage difference between desired and initial balances is made up. Of course, if the adjustment coefficient γ equals unity, (5b) is reduced to (5a). The theoretical foundation of (5b) in the context of demands for liquid assets is unclear. Stephen Goldfeld discusses other partial adjustment mechanisms but their theoretical underpinnings are in our minds hardly more satisfactory. In any case, by assuming (5b) we can compare a limited number of regressions with numerous studies, being sure that any difference in results arise from other features of the model.

Consider the simpler instantaneous adjustment case first. Substituting (5a) and (4) in (1) gives

$$(6) \quad M_t = \alpha_0 + \alpha_1(1 - \lambda\beta)y_t + \alpha_1\lambda\beta(1 - \beta) \sum_{i=0}^{\infty} \beta^i y_{t-1-i} + \alpha_2 r_t$$

---

[3] The slight modification of Friedman's measure facilitates computations since we assume money demand is linear in the logarithms (constant elasticity). Friedman's weights sum to slightly more than one to allow for growth; we could allow for growth by introducing a constant in (2), but this would not change the form of the regressions or their interpretation.

[4] Henceforth, expected income without modification will always refer to short-run expected income. Normal or permanent refers to long-run expected.

[5] Given consistent realization of income expectations, i.e., $y_t^e = y_{t+1}$, it can be shown that actual and normal income converge.

Since all variables are expressed in logarithm, α_1 and α_2 are elasticities. Previous empirical studies suggest that α_1, the elasticity of real balances with respect to income should be about unity, and α_2, the interest rate elasticity, should be between zero and one. The elasticity should be greater for long rates than short rates but within the stated limits in either case. Equation (6) cannot be estimated as it stands. Since the regression coefficients beginning in $t-1$ are restricted to fall exponentially, apply a Koyck transformation. Lagging (6), multiplying both sides by β, and subtracting gives a reduced equation which, given the empirical work that follows, is most conveniently written as

$$\text{(7a)} \quad M_t - \beta M_{t-1} = C_0 + C_1 y_t + C_2 y_{t-1} + C_3[r_t - \beta r_{t-1}]$$

where

$$C_0 = (1-\beta)\alpha_0$$
$$C_1 = \alpha_1(1-\lambda\beta)$$
$$C_2 = \alpha_1\beta(\lambda-1)$$
$$C_3 = \alpha_2$$

The behavioral parameters in terms of β and the coefficients of the reduced equation are

$$\text{(8)} \quad \lambda = (C_1\beta + C_2)/\beta(C_1 + C_2)$$
$$\alpha_1 = (C_1 + C_2)/1 - \beta$$
$$\alpha_2 = C_3$$

Examination of (7a) and (8) shows that if lagged balances and current and lagged interest rates were included as independent variables, the structural parameters α_2 and β would be overidentified. In the empirical work we estimate (7a) by iterating over β.[6] Use of the Koyck transformation produces another problem in addition to overidentification, namely the introduction of lagged balances. This problem is not avoided by a simple transformation that includes lagged balances on the left-hand side of (7a). We take this problem up again in the empirical Section III.

If only partial adjustment occurs during the period, then using (5b) instead of (5a) and following procedures similar to those which yielded (7a) gives

$$\text{(7b)} \quad M_t - \beta M_{t-1} = C_0' + C_1'(M_{t-1} - \beta M_{t-2}) + C_2' y_t + C_3' y_{t-1} + C_4'(r_t - \beta r_{t-1})$$

Equation (7b) is another form of the equations to be estimated. Again if M_t were the dependent variable and M_{t-1} and M_{t-2} as well as r_t and r_{t-1} were included as separate independent variables, the structural parameters would be overidentified. To save space we simply assert that as was true for (7a), the C_i' coefficients in (7b) are functions of the behavioral parameters and can be unscrambled to determine the behavioral coefficients, given β.

Goldfeld and David Laidler (1969) state that the transactions approach focuses on income, and presumably expected income, as the scale variable; the asset approach uses long-run variables such as permanent income or wealth. Given this view, our model indicates that λ is the crucial parameter in the transactions-asset motive controversy. Recall that if $\lambda = 0$, expected income reduces to current income, and the transactions/finance approach is supported. It is unlikely that $\lambda=0$ since Laidler (1966a) showed that a distributed lag of income is superior to current income. The asset approach to money demand presumably is valid if $\lambda=1$ as expected income then equals permanent income as computed by Friedman. If λ is significantly different from unity (but not necessarily equal to zero), then according to our model money demand depends on expected income or transactions and not on permanent income or assets. Of course, λ may be significantly different from unity and the asset motive may still predominate simply because the Friedman method of computing permanent income contained in the model is incorrect, the correct method of computing permanent income being our modified exponentially weighted average. If we relied exclusively on doctrinal history, the equation discriminates between asset and transactions only given the Keynes-Friedman assumption. This lack of definiteness is largely eliminated by considering optimal prediction criteria in a stochastic model and by the relationship between coefficient estimates and the definition of money in the empirical tests.

---

[6] It may appear that our estimation technique is arbitrary; we could have searched over α_2. However, given the objective of minimizing the squared residuals, which as we argue later is the correct objective, it can be shown that the direction of search is irrelevant, assuming perfect searches. In other words, searching for values of α_2 and estimating λ, β, and α_1 would yield the same parameters as those shown in Table 1. We searched over β simply because the theory suggested a range of values for this parameter. While we also could have scanned over β, computed permanent income, and estimated equation (6) directly, this alternative is less desirable on several counts. We would have to use estimates of *NNP* for earlier periods (pre-1897), and our impression is that the estimates are less reliable the further one goes back in time. As β approaches unity, data sufficient to compute permanent income are not available. War year data are completely excluded in (7a) and (7b) but are necessary in (6) to compute normal income.

II. Autoregressive Processes and Optimal Predictors

Suppose that the myriad of factors influencing income are such that its time-series can be regarded as generated by independent "shocks" a_t. As a specific example, let income be a linear function of independently and normally distributed random variables with mean zero and common variance where the coefficients or weights of the lagged variables are equal,

$$y_t = a_t + k_0 \sum_{i=1}^{\infty} a_{t-i} \tag{9}$$

where $0<k_0<1$

To repeat, the a's are independent, but looking at (9) shows that output y_t will be serially correlated. Muth has analyzed this stochastic model, showing that permanent income as computed by Friedman is the optimal (in the sense of minimizing the mean square error of prediction) predictor of income in any future time period, given simply the observed past incomes. A more general stochastic model than Muth's is

$$y_t = a_t + \sum_{i=1}^{\infty} (k_0 + (1-k_0)k_1^i)a_{t-i} \tag{10}$$

where k_0 and k_1 lie between zero and one. Throughout most of this section, we assume the parameters are known. Our model reduces to Muth's when $k_1=0$ so that the weights of past disturbances are all equal to k_0. As Muth notes, equation (9) implies that the shocks contain two distinct effects, permanent and temporary. In the current time period the shock has a temporary effect $(1-k_0)$ plus a permanent effect k_0, so that current shocks have a weight of unity. The temporary effect is zero in all subsequent time periods. According to (10) the shock has a "lingering" and permanent effect, given $k_1\neq 0$. The weight of a shock is unity in the current period and in subsequent periods approaches k_0, its permanent effect. It certainly is reasonable, but contrary to the Friedman-Muth model, to assume that a shock in, say, 1973 has a greater effect on output in 1975 than 1985. An observation generated by (10) may also be expressed as an infinite weighted sum of past observations plus a random variable. Repeatedly lagging (10) and substituting yields

$$y_t = w_1' y_{t-1} + \sum_{i=0}^{\infty} w_{1+i} y_{t-1-i} + a_t \tag{11}$$

where
$$w_1' = k_1/(1-k_0(1-k_1))$$
$$w_{1+i} = k_0(1-k_0)(1-k_1)^2 \cdot (1-k_0(1-k_1))^{i-1}$$

Now suppose that at period t we are to

forecast y_{t+s} which is to be a linear function of y_t and previous observations. Since we are interested in forecasts with different lead times, denote the optimal forecast at t for $s \geq 1$ periods forward by $y_t^e(s)$. In terms of the notation of the previous section $y_t^e = y_t^e(1)$. How does one weight the observations in order to minimize the mean square error of prediction? Let the optimal forecast be

$$\text{(12)} \qquad y_t^e(s) = v_1(s)y_t + v_2(s)y_{t-1} + v_3(s)y_{t-2} + \dots$$

where the weights $v_1(s)$, $v_2(s)$, . . . , to be determined are generally functions of the lead time s of the forecast. Clearly from (11) actual income in period t+s is

$$\text{(13)} \quad y_{t+s} = w_1' y_{t+s-1} + \sum_{i=0}^{\infty} w_{1+i} y_{t+s-1-i} + a_{t+s}$$

The mean square error of prediction at period t is $E[y_{t+s} - y_t^e(s)]^2$. It is well known that the variance is less than the second moment about any number other than the mean. Consequently, since the prediction error is minimized given the realized shocks through period t, the optimal predictor is the conditional expectation.[7]

$$\text{(14)} \qquad y_t^e(s) = E[y_{t+s} \mid y_t, y_{t-1}, \dots] = E[y_{t+s} \mid t]$$

Examination of (10) and (11) shows that y is not time stationary. Let $s=1$ and take the (conditional) expectation of (13)

$$\text{(15)} \qquad E[y_{t+1} \mid t] = w_1' y_t + \sum_{i=0}^{\infty} w_{1+i} y_{t-i}$$

Equating coefficients in (12) and (15) gives

$$\text{(16)} \qquad \begin{aligned} v_1(1) &= w_1' + w_1 \\ v_i(1) &= w_i \qquad i = 2, 3, \dots \end{aligned}$$

For $s=2$ and making use of (15), the conditional expectation of (13) is

$$\text{(17)} \quad \begin{aligned} &E[y_{t+2} \mid t] \\ &= (w_1' + w_1)E[y_{t+1} \mid t] + \sum_{i=1}^{\infty} w_{1+i} y_{t+1-i} \\ &= [(w_1' + w_1)^2 + w_2]y_t + \sum_{i=1}^{\infty} [(w_1' + w_1)w_{1+i} + w_{2+i}]y_{t-i} \end{aligned}$$

Again equating coefficients gives

$$\text{(18)} \quad \begin{aligned} v_1(2) &= (w_1' + w_1)^2 + w_2 \\ &= [v_1(1)]^2 + w_2 \\ v_i(2) &= (w_1' + w_1)w_i + w_{1+i} \\ &= v_1(1)w_i + w_{1+i} \qquad i = 2, 3, \dots \end{aligned}$$

Proceeding in this manner it can be shown that in terms of the original parameters k_0 and k_1,[8]

$$\text{(12}'\text{)} \qquad y_t^e(s) = [1 - \theta(s)]y_t + \theta(s)y_t^n$$

where

$$\theta(s) = \frac{(1 - k_0)(1 - k_1^s)}{1 - k_0(1 - k_1)}$$

$$y_t^n = k_0(1 - k_1) \sum_{i=0}^{\infty} (1 - k_0(1 - k_1))^i y_{t-i}$$

The optimal forecast is a modified exponentially weighted average, that is, a weighted average of the current and normal values at period t. The weights vary according to the lead time of prediction in our model with lingering and permanent shocks. As we expected if $k_1=0$, then

---

[7] For a rigorous proof, see G. E. P. Box and J. M. Jenkins, p. 128.

[8] While there are some advantages in letting the coefficient of y_{t-1} in (15) be $w_1' + w_1$, at this point a transformation is useful. Let $Z_1 = w_1' + w_1$ and $Z_i = w_i$; $i = 2, 3, \dots$. Box and Jenkins, p. 142, showed that

$$v_i(s) = Z_{i+s-1} + \sum_{h=1}^{s-1} Z_h v_i(s - h)$$

Using this equation and the well-known recursion formula for permanent income, (12′) follows more readily.

$\theta(s)=1$ and normal income is the optimal forecast for all periods. Contrary to the Muth-Friedman model given by (9), our optimal predictor with $k_1 \neq 0$ depends on the time interval of the forecast.

Equation (12′) yields optimal forecasts of income given that the forecasts are a linear function of y_t and previous observations alone. Could the forecasts be improved by incorporating further information and behavioral and accounting relationships? The answer is not clearly yes. For example, Charles Nelson showed that beyond the sample period over which the FRB-MIT-PENN model was fitted the mean square error was more than twice as large as the errors in predicting income with a single autoregressive equation, presumably because of the specification and sampling errors in large scale models. Treating income "as if" it were generated by an autoregressive process and estimating its parameters certainly facilitates forecasting; whatever other data besides lagged income one has available may be superfluous or even dangerous.

It is desirable to reconsider forecasting one period ahead and in the distant future. For s=1, equation (12′) becomes after some rearranging

$$(19)\quad y_t^e(1) = [1-(1-k_0)(1-k_1)]y_t + [(1-k_0)(1-k_1)]y_{t-1}^n$$

The doctrinal forecast of income given by (4) and the optimal forecast derived from the time-series model (19) have the same form where

$$(20)\quad \lambda = (1-k_0)(1-k_1)/1-k_0(1-k_1)$$
$$\beta = 1-k_0(1-k_1)$$

Is Friedman's normal or permanent measure an optimal forecast in a model where optimal forecasts for all periods are not identical? No direct and definitive answer to this question is possible since Friedman purposely left vague the theoretical concept of permanent income. An indirect approach suggests itself. Calculate expected income for alternative lead times and see if some expected income approaches Friedman's empirical construct. Given our restrictions on the values of k_0 and k_1, examination of (12′) shows that θ is strictly positive. Differentiating with respect to the lead time

$$(21)\quad \partial\theta/\partial s = \frac{-(1-k_0)k_1^s \log k_1}{1-k_0(1-k_1)} > 0$$

The weight on normal income is positively related to the lead time of the forecast. Expressed in terms of λ and β, which are estimated in the following section, equations (12′) and (20) show that as the lead time becomes large the forecast approaches

$$(22)\quad y_t^e(\infty) = \left[\frac{(1-\lambda)(1-\beta)}{1-\beta(1-\lambda)}\right]y_t + \left[\frac{\lambda}{1-\beta(1-\lambda)}\right]y_t^n$$

Anticipating the empirical results, using the narrow definition of money (currency plus demand deposits), the minimum and maximum estimated values of $\lambda/[1-\beta(1-\lambda)]$, are .91 and .96, respectively, in regressions using alternative interest rates and making alternative assumptions about the adjustment of desired to actual money balance. For $y_t^e(1)$ the weight of y_t^n is λ, which varies between .56 and .77. Thus our empirical results indicate the optimal forecast of income in the distant future is approximately given by Friedman's normal income measure.

$$(23)\quad y_t^e(\infty) \cong y_t^n$$

While other interpretations are possible, equating permanent income with a long-run forecast is a reasonable interpretation of Friedman and assures optimality for his income measure.

III. Estimates of the Model

In this section we present regression estimates of the model developed in Section I using annual data covering the period 1897–1960 with the war years 1917–19 and 1941–45 excluded. The data which are expressed in *logs* include: (a) National Bureau of Economic Research (*NBER*) unpublished figures on *NNP*, the four-six month prime commercial paper rate, and an implicit price index (1929 base); (b) M^1, currency and demand deposits, and the yield on twenty-year corporate bonds provided by Gregory Chow up to 1958 and the Federal Reserve *Bulletin* and *Statistical Abstract* for 1959–60; and (c) M^2, currency plus demand deposits plus time deposits, taken from Friedman and Schwartz (1963a). The data and time periods were chosen for a basis of comparison with other empirical studies on the demand for money such as Chow (1966, 1968), Friedman (1959), Laidler (1966a,b), and Allan Meltzer. The use of annual data in no way implies that money balances finance outlays one year hence.

Although we emphasize the role of money as a medium of exchange, we do not simply drag M^2 along to multiply the number of regressions. Our model implies that the values of some parameters depend crucially on the definition of money. For example, λ, a parameter discriminating between asset and transactions motives, should be closer to unity when money is defined as M^2 as opposed to M^1. Desired medium of exchange balances depend on planned short-term transactions, which gives a value of λ different from unity. The broader definition of money includes an asset which is not a medium of exchange. For such an asset, λ should be approximately unity since total assets or permanent income is a better determinant than planned transactions, as Keynes noted (1930, p. 36). Since time deposits were approximately 40 percent of M^2, in the M^2 regressions λ may not be significantly different from unity. The appropriate income constraint depends on the empirical definition of money.

To our best knowledge, studies that include permanent income or nonhuman wealth and ignore the transactions (finance) motive find that M^1 is systematically more sensitive to market interest rates than M^2. That is, for both short- and long-term interest rates, the absolute elasticity is substantially greater for the narrower definition of money. Only H. R. Heller finds that the interest rate elasticity is greater for M^2, and his preferred constraint variable is nominal income. From the theory of omitted variables, the systematic relationship between estimated interest rate elasticities and the definition of money is as it should be when the transactions motive is ignored. The asset, general theory of choice approach is silent about the relative interest rate elasticities of different assets. The percentage shifts from currency and demand deposits to all other assets including time deposits in response to higher relative interest rates is as a first approximation, according to this view, equal to the percentage shift from time deposits to other assets. As market interest rates rise, the time deposit rate generally fails to increase proportionally so that the relative time rate falls. Moreover, the relative rate on M^1 falls even more because currency bears no interest and interest payments on demand deposits have been prohibited since the 1930's. Services which increase the return on demand deposits cannot be expanded rapidly. Lack of a good series on deposit rates forces their omission. Consequently, for a given increase in market rates, there will be a greater shift from demand than time deposits to all other assets, and some demand deposits shift to time deposits as the relative time/demand deposit rate rises. The estimated market interest rate elasticity should be greater for the narrower definition of money given the crucial assumption that the elasticity with respect to interest rates in general is the same for both assets. In other words, if the "true" elasticities of M^1 and M^2 with respect to market rates are equal, the omission of deposit rates will bias downward (towards zero) less the estimated elasticity of M^1.

There should be no systematic relationship between the estimated market rate elasticities and the definition of money if the transactions motive significantly explains M^1. Transactions balances are highly interest inelastic according to classical and Keynesian views, or at least more inelastic than other assets like time deposits. The omission of deposit rates biases towards zero the coefficient of market rates less for M^1 than M^2, but the true coefficient is absolutely smaller in the M^1 regressions. Therefore, the estimated market rate elasticities may even be smaller for M^1. It is the pattern of interest rate coefficients which sheds light on the problem previously posed: do the distributed lag (4) and our reduced equations (7a) and (7b) really contain a transactions element or are they simply providing better measures of an asset constraint?

The ordinary least squares (*OLS*) estimates of (7a) and (7b) are presented in Table 1. Since the coefficients of the reduced equations are of little interest, we only present the structural parameters. The standard errors of the estimated structural parameters, which are nonlinear functions of the reduced parameters, were obtained using a formula presented by Lawrence Klein. The table shows the standard errors in parentheses. β is the parameter used to compute permanent income had we not applied a Koyck transformation to get (7a) from (6). In estimating (7a) and (7b) we scanned over β at 0.05 intervals between zero and one and chose that value of β which minimizes the standard error of the estimate. Since the dependent variable clearly depends on the value of β, minimizing the unexplained

TABLE 1—ESTIMATES OF MONEY DEMAND FUNCTIONS
Equations (7a) and (7b)

Dependent Variable	β	Transactions/ Assets λ	Income Elasticity α_1	Interest Rate Elasticity		Stock Adjustment Coefficient γ	$\frac{\lambda}{1-\beta(1-\lambda)}$	Unbiased Squared Standard Error of the Estimate
				α_2^S	α_2^L			
M^1	.850	.611	.958	−.057			.913	.0019
		(.082)	(.073)	(.024)				(1.57)
M^2	.750	.840	1.127	−.093			.955	.0008
		(.079)	(.035)	(.015)				(2.69)
M^1	.900	.722	.886		−.310		.963	.0016
		(.110)	(.105)		(.084)			(1.06)
M^2	.850	.878	1.065		−.265		.978	.0009
		(.094)	(.053)		(.069)			(1.63)
M^1	.900	.563	.909	−.052		.972	.928	.0018
		(.126)	(.152)	(.021)		(.064)		(1.06)
M^2	.700	.854	1.191	−.111		.896	.951	.0008
		(.082)	(.052)	(.016)		(.055)		(3.27)
M^1	.850	.765	.937		−.354	.992	.956	.0015
		(.105)	(.069)		(.092)	(.061)		(1.57)
M^2	.850	.877	1.077		−.263	1.041	.979	.0010
		(.068)	(.077)		(.067)	(.057)		(1.63)

sum of squares is preferable to maximizing the *relative* explained sum of squares, R^2. The number in parentheses in the last column is the mean of the dependent variable.

We generally limit discussion of the table to the first four regressions, which assume instantaneous adjustment. Looking at the latter stock-adjustment regressions we see that γ is virtually equal to unity, indicating rapid adjustment of actual to desired balances. Moreover, the value of no other parameter is sensitive to the adjustment assumption made. The estimates of γ are extremely encouraging. If the most liquid asset money, however defined, is not virtually completely adjusted to its desired level within a year, how long are other lags? While Feige also found complete adjustment within a year, Chow (1966) and Martin Bronfenbrenner and Thomas Mayer have found quite long lags.

The income elasticities are in accord with the study by Meltzer. They are close to unity, the only significant difference being the regression of M^2 on the short-term interest rate. The income elasticity of M^2 is uniformly about 0.15 greater than the elasticity of M^1 in corresponding equations. Like Laidler (1969a), we find no support for Friedman's (1959) view that money is a luxury.

The values of λ are entirely consistent with our hypothesis. Given the doctrinal and time-series justifications for considering the modified exponentially weighted average as expected income, and equating approaches to money demand with the time horizon implied by the best scale variable, finance requirements and not assets explain money demand particularly when money is defined as M^1. A value of λ significantly different from unity in the M^1 regressions indicates that short-period expected income is the determinant of the medium of exchange, and short-period expected income is different from Friedman's permanent income. When one broadens the definition of money to include time deposits, permanent income plays a more important role. The value of λ rises in the M^2 regressions and while still less than unity, the differences are of only marginal significance at the 5 percent level in the regressions with the short-term interest rate. Expected income does not equal current income; λ is significantly different from zero. Planned expenditures or the finance motive rather than current expenditures determine money demand. The tendency to equate the superior performance of permanent over current income in long-run or instantaneous adjustment regressions like (1)–(4) with superiority of the asset motive is incorrect. The (planned) transactions motive is significant, but to capture it (7a) shows that one must include more in the regression than current income and current interest rates.

The interest rate elasticities are relatively low but significant, at least at the 5 percent level. From a purely technical viewpoint, the lower elasticities for the short rate may simply reflect the larger swings of short rates over long rates. As we expected there is no systematic relationship between interest elasticities and the definition of money. M^1 rather than M^2 is less responsive to movements in the short-term rate while the reverse holds for the long rate.

The next to last column in Table 1 shows the weights of permanent income in distant forecasts in an optimal forecasting model. Ordinary least squares (*OLS*) estimates of (7a) are acceptable. If the disturbance terms follow the classical least squares properties, J. Johnston shows that with a lagged dependent variable *OLS* estimators are biased, but the bias in our case is quite small since the sample size is more than fifty. Moreover, the *OLS* estimators are maximum-likelihood, consistent, and asymptotically efficient. If the disturbance term is autocorrelated, however, *OLS* is no longer even consistent and another estimation technique would be necessary. The Durbin-Watson statistic is biased towards indicating random disturbances in the presence of the lagged dependent variable on either left- or right-hand side of the equation. Therefore, to test for autocorrelation according to Johnston (p. 313) we estimate (7a) by *OLS* with lagged balances on the right-hand side and computed Durbin's *h*-statistic. Using both definitions of money and both interest rates, in no case could we reject at the 5 percent level the null hypothesis that the disturbance terms were uncorrelated. Consequently, nothing more complicated than the *OLS* estimators in Table 1 are necessary.[9]

How does our money demand function compare with other estimates? Having already shown that our coefficients are "reasonable," all that remains is a comparison of the fit. Since equations may be subject to substantial sample bias, we feel evaluating predictive power beyond the period used to estimate the regression is far superior to comparing R^2 or errors in the sample period. Evaluating the predictive ability of our equations against all others using alternative combinations of interest rates and money definitions would, of course, be impossible. Money demand regressions are typically log-linear, and the narrow definition of real balances and the long-term rate of interest are more commonly used. To test various hypotheses, Chow (1966, 1968) and Laidler (1966a,b) estimated a battery of regressions using as independent variables current and lagged values of actual income, Friedman's permanent and temporary income, wealth, interest rates, and lagged balances. Both Chow and Laidler find that permanent income as computed by Friedman is the appropriate constraint for a long-run money demand function, and Chow further includes current income and lagged balances in his short-run function to capture a modified stock-adjustment mechanism. Therefore, we compare the errors in predicting the *log* of real balances with our (7a), the parameters given by the third line of Table 1, against the permanent income variant of equation (7a), (assumes $\lambda = 1$), and the following equations:[10]

$$M_t^1 = -0.989 - 0.721 r_t^L - 0.145 N_t + 1.145 y_t^P$$

Laidler (1966b, p. 522)

$$M_t^1 = 0.281 - 0.298 r_t^L + 0.216 y_t^P + 0.162 y_t + 0.607 M_{t-1}^1$$

Chow (1968, p. 1242)

where N is population, y_t^P is Friedman's permanent measure, and the other variables are as previously defined.

Referring to Table 2, the dates under the authors' names are the estimation period. Prediction errors for five years beyond the estimation period and for 1961–63 are calculated. The first line shows the errors using our (7a), while the second line shows the errors for the permanent income variant of (7a). That is, we

---

[9] If one assumed that (1) contained a normally and independently distributed error term, then the disturbance in (7a) would be serially correlated. In this case the standard errors presented in Table 1 are biased and the estimation technique suggested by J. Phillip Cooper would be appropriate. One is never on firm ground in specifying disturbance terms, and the assumption of the first sentence is no more reasonable than the null hypothesis in the text.

[10] We ignore Meltzer's important study which emphasizes nonhuman wealth because the exchange between Chow and Karl Brunner and Meltzer showed that "... some very slight differences in the measure of wealth that was used" (p. 1237) can alter coefficients dramatically. The aggregate equivalent of Laidler's equation which is in per capita terms is given. That is, Laidler's equation is $M^1/N = -0.989 - 0.721\, r_t^L + 1.145\, (y_t^P/N)$, using our notation where M^1 is *log* real balances, and similarly for other variables.

TABLE 2—PREDICTION ERRORS

		Prediction Errors for *Log* (Narrow) Real Balances			
Study	Independent Variables	Period	Mean Error	Mean Absolute Error	Standard Deviation of Errors
Meyer-Neri (1897–1960)	y_t, y_{t-1}, M_{t-1}, r_t, r_{t-1} (constrained)	1961–65	−.009	.014	.019
		1961–63	.001	.015	.016
Meyer-Neri (1897–1960)	y_t, M_{t-1}, r_t r_{t-1} (constrained)	1961–65	−.006	.023	.029
		1961–63	−.020	.030	.031
Laidler (1919–60)	y_t^P, r_t, N	1961–65	.030	.030	.028
		1961–63	.031	.031	.016
Chow (1897–1958)	y_t^P, y_t, r_t, M_{t-1}	1959–63	.053	.053	.020
		1961–63	.041	.041	.014
		1959–61	.058	.058	.015

assumed $\lambda=1$, deleted y_{t-1} as an independent variable, and then iterated over β at 0.05 intervals so that the ratio of the coefficient of r_t and r_{t-1} is constrained to be minus β. The value of β minimizing the squared standard error of the regression was chosen. With regard to the transactions/finance and wealth controversy, the first two lines are most relevant since the same period and estimation method was used. Our (7a) does appreciably better than the permanent income variant however the prediction error is measured. The mean error in predicting *log* money balances is virtually equivalent to the relative error in predicting the level of money balances.[11] The basic hypothesis underlying lines 2 and 3—money demand depends on permanent income and the long rate—is identical, the difference being the sample period and estimation technique.[12] Computing permanent income directly rather than applying a Koyck transformation yields systematic underestimation of real balances, the mean and mean absolute errors are equal, but has little effect on the standard deviation of the errors for the full 1961–65 period. Chow's short-run equation also consistently underestimates money balances. In terms of the standard deviation of the prediction errors, there is little difference between our (7a) and Chow's equation.[13] If we had chosen to compare root mean square errors, the results would be more favorable for our hypothesis since our mean prediction error is small.[14] The greater variability of the errors for 1958–63 relative to the subperiods might be explained by the autocorrelation of Chow's regression (Durbin's h-statistic is significant at the 1 percent level). Table 2 does not demonstrate beyond doubt the superiority of the transactions/finance approach; it does show that this approach does at least as well as the wealth approach. Moreover, we find our estimates of complete adjustment of actual to desired balanced within a year more plausible than Chow's long lags.

---

[11] $Log M - log \hat{M} = -log \quad (\hat{M}/M) \cong -((\hat{M}/M)-1) = (M-\hat{M})/M$, since $\hat{M}/M$ is close to unity.

[12] We searched to find the optimal income weights in the money demand function while Friedman's permanent income has weights which maximize the correlation between income and consumption. Even if the income weights allowing for growth and the sample period of the dependent variable were identical, applying a Koyck transformation instead of computing normal income directly could lead to different estimates because a longer income series is necessary to compute normal income directly and the unobserved errors of the regressions would not be identical.

[13] Some caution is necessary in comparing Chow's prediction errors against the others because a slightly different variable is being predicted. Chow deflates nominal balances by the personal consumption expenditures price index while Laidler and the authors deflate by the *NBER* implicit price index for *NNP*, variant III.

[14] The mean square error, which probably is the most important property of an estimator, can be computed from Table 2 by noting that it equals the variance of the estimator plus the square of the mean error (bias).

IV. Summary

The appropriate definition of money is the subject of substantial controversy. Nevertheless, everyone is willing to include currency and demand deposits. Particularly if money is defined to include only these two assets, it is inconceivable in our minds that recognizing that money is a medium of exchange will add nothing to the empirical estimates of money demand. If at this point we may engage in some armchair empiricism and philosophizing, the quest for generality in economics has gone too far. Laidler (1969) wonders why anyone would consider motives for money when motives for refrigerators are ignored. Let us admit it, refrigerators do keep beer and watermelon cold, and this may improve estimates of the demand for refrigerators. A less than exhaustive but more than cursory examination of our newspapers shows that small refrigerators (less than 5 cubic feet) are much more heavily advertised (and sold) in the spring and summer than in the winter. Money is the medium of exchange; this bit of information should add something. It does. In simple regressions of money on an interest rate and either current or permanent income, the latter has greater significance. This has led to the incorrect conclusion that the asset approach dominates. The amount of real balances demanded depends on real planned transactions, which are not well measured by either current or permanent income alone. The Keynes-Friedman approach leads to money demand being a function of a modified exponentially weighted average of income, which measures finance requirements.

The pattern of interest rate coefficients further supports the transactions/finance approach. While this approach is contrary to Friedman's views, we find no support for a Keynesian liquidity trap. We ran several tests for a liquidity trap and like Laidler (1966b) could find none. Although statistically significant, the interest rate elasticity of money is very low, about one-half other estimates. The empirical results described in this paper suggest the desirability of further work along the lines of the recent theoretical literature emphasizing that money reduces transactions and information costs and saves time.

REFERENCES

G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control,* San Francisco 1970.

M. Bronfenbrenner and T. Mayer, "Liquidity Functions in the American Economy," *Econometrica,* Oct. 1960, *28,* 810–34.

K. Brunner and A. Meltzer, "Comment on the Long-Run and Short-Run Demand for Money," *J. Polit. Econ.,* Nov./Dec. 1968, *76,* 1234–40.

G. Chow, "On the Long-Run and Short-Run Demand for Money," *J. Polit. Econ.,* Apr. 1966, *74,* 111–31.

———, "Long-Run and Short-Run Demand for Money: Reply and Further Note," *J. Polit. Econ.,* Nov./Dec. 1968, *76,* 1240–43.

J. P. Cooper, "Asymptotic Covariance Matrix of Procedures for Linear Regression in the Presence of First-Order Autoregression Disturbances," *Econometrica,* Mar. 1972, *40,* 305–10.

D. Dutton and W. Gramm, "Transactions Costs, the Wage Rate, and the Demand for Money," *Amer. Econ. Rev.,* Sept. 1973, *63,* 652–65.

E. L. Feige, "Expectations and Adjustments in the Monetary Sector," *Amer. Econ. Rev. Proc.*, May 1967, *57*, 462–73.

M. Friedman, "The Demand for Money: Some Theoretical and Empirical Results," *J. Polit. Econ.*, Aug. 1959, *67*, 327–51.

———, *A Theory of the Consumption Function*, New York 1957.

———, "A Theoretical Framework for Monetary Analysis," *J. Polit. Econ.*, Mar./Apr. 1970, *78*, 193–238.

——— and A. Schwartz, (1963a) *A Monetary History of the United States, 1867–1960*, Princeton 1963.

——— and ———, (1963b) "Money and the Business Cycle," *Rev. Econ. Statist.*, suppl., Feb. 1963, *45*, 32–64.

S. Goldfeld, "The Demand for Money Revisited," *Brookings Papers, 3*, Washington 1973, 577–638.

H. R. Heller, "The Demand for Money: The Evidence from the Short-Run Data," *Quart. J. Econ.*, May 1965, *79*, 291–303.

J. Johnston, *Econometric Methods*, 2d ed., New York 1972.

J. M. Keynes, *A Treatise on Money*, Vol. I, London 1930.

———, *The General Theory of Income, Employment, and Money*, New York 1936.

———, "The 'Ex-Ante' Theory of the Rate of Interest," *Econ. J.*, Dec. 1937, *47*, 663–69.

L. R. Klein, *A Textbook of Econometrics*, Evanston 1953.

D. Laidler, (1966a) "Some Evidence on the Demand for Money," *J. Polit. Econ.*, Feb. 1966, *74*, 55–68.

———, (1966b) "The Interest Rate and the Demand for Money—Some Empirical Evidence," *J. Polit. Econ.*, Dec. 1966, *74*, 543–55.

———, *The Demand for Money: Theories and Evidence*, Scranton 1969.

A. Meltzer, "The Demand for Money: The Evidence from the Time Series," *J. Polit. Econ.*, June 1963, *71*, 219–46.

F. Modigliani and R. Sutch, "Innovations in Interest Rate Policy," *Amer. Econ. Rev. Proc.*, May 1966, *56*, 178–97.

J. F. Muth, "Optimal Properties of Exponentially Weighted Forecasts of Time-Series with Permanent and Transitory Components," *J. Amer. Statist. Assn.*, June 1960, *55*, 297–306.

C. R. Nelson, "The Prediction Performance of the FRB-MIT-PENN Model of the U.S. Economy," *Amer. Econ. Rev.*, Dec. 1972, *62*, 902–17.

G. L. S. Shackle, "Keynes and Today's Establishment in Economic Theory: A View," *J. Econ. Lit.*, June 1973, *11*, 516–19.

Board of Governors of the Federal Reserve System, "A New Measure of the Money Supply," *Fed. Res. Bull.*, Oct. 1960, *46*, 1102–23.

———, *Fed. Res. Bull.*, July 1961, *47*, p. 799.

U.S. Bureau of the Census, *Statistical Abstract of the United States*, Washington 1959–60.

Short-Run Asset Effects on Household Saving and Consumption: The Cross-Section Evidence

By IRWIN FRIEND AND CHARLES LIEBERMAN*

This paper introduces a new approach to the measurement of the short-run effects of capital gains on household saving and consumption. Previously, the timing and magnitude of these asset effects have been estimated exclusively from time-series data—either from a single equation consumption function or, more commonly in recent years, from a large-scale econometric model. The basic difficulty with the approach based on aggregate time-series data is the small number of independent observations available for determining the appropriate parameters, especially in a model with a large number of structural equations and a large number of alternative forms for each equation.

The economic importance of obtaining further insights into the short-run impact of capital gains on saving is highlighted by the MIT-PENN-SSRC (*MPS*) model, which asserts that one of the most important channels of transmission of monetary policy is the wealth effect on consumption.[1] The updated (January 1973) version of the *MPS* model implies that a \$1.00 capital gain increases the long-run rate of consumption by \$.054, with most of this effect taking place in the first year. However, while the estimated long-run effect has changed only moderately in different recent versions of the model, the timing of that effect has changed greatly. Moreover, as discussed subsequently, other time-series analyses have provided radically different results for the magnitude and timing of asset effects on consumption.

Given the large margin of indeterminacy in estimating the pattern of asset effects on consumption from time-series data, it is surprising that survey data have not been used as a check on the time-series analysis. This is not to say that survey data do not have deficiencies of their own, but simply that the paucity of independent observations characterizing time-series, compounded by temporal changes in the economic structure, cannot be remedied without the use of cross-section information. The most serious deficiencies in survey or cross-section data can be corrected with the availability of sufficient resources. Panel data covering the same households (or groups of households) over several consecutive periods of time are necessary to avoid some of the limitations of a single cross-section, but highly useful information on short-run asset effects can be obtained from appropriate information collected in a single cross-section.

The cross-section data used for this paper consist of information on saving, income, and assets from two surveys of 2,159 households in 1963 (with income

---

* Professor of finance and economics, University of Pennsylvania, and assistant professor of economics, University of Maryland, respectively. We wish to thank the National Science Foundation for financial support and the Board of Governors of the Federal Reserve System for making previously unpublished survey data available to us.

[1] The latest published version of the *MPS* model appears in Franco Modigliani, but a more recent version is available in *MIT-PENN-SSRC Econometric Model of the United States*, January 1973.

data for 1962 as well as 1963) conducted by the Federal Reserve Board (*FRB*).[2] The surveys, which oversampled upper income groups, collected detailed information not only for income, saving, and the values of major categories of assets, but also for the amounts of individual stocks held at the beginning of the year. For purposes of this paper, we were able for the first time to use the data on individual stockholdings to estimate capital gains (largely unrealized) on a major and most volatile part of household wealth. As a result, the effect of capital gains on saving during the year can be studied on the basis of a large number rather than a handful of observations.

Our analysis of the survey data indicates that household saving in 1963 was reduced by about \$.02 to \$.04 for every dollar of capital gains on corporate stock during that year, with the precise figure obtained depending on the form of the regression utilized and the nature of the adjustments made to the original *FRB* sample. The bulk of our estimates, which are concentrated in the lower half of the .02 to .04 range, are above the corresponding figures implied by John Arena (1964, 1965), Kul Bhatia, Martin Feldstein, and Feldstein and George Fane, and modestly less than the figure implied by the latest version of the *MPS* model.

While estimates of long-run asset effects on household saving are also provided by our analysis of the 1963 survey data, they are not as stable as the capital gains coefficients. Moreover, as is well known, the long-run asset coefficients in saving regressions based on household survey data are downward biased in absolute value since assets serve as a proxy for savings tastes as well as a measure of resources available for consumption and investment. To obtain definitive insights into the timing and magnitude of asset effects on consumption, it will probably be necessary to collect new panel data on household savings, income, assets, and capital gains. However, even with the cross-section information already available, it is possible to improve on our present knowledge of such effects.

I. Earlier Work

The only extensive use of cross-section data in earlier studies of asset effects on household savings and consumption consisted of the estimation of long-run effects. These effects were generally rather small, even when they had the theoretically expected sign and were statistically significant.[3] They are understated because of differences in household saving tastes which cannot adequately be kept constant in an analysis of single cross-section data.

Earlier time-series studies by Arena (1964, 1965), Bhatia, Feldstein, and Feldstein and Fane included estimates of both long-run and short-run asset effects. All reported either small or statistically insignificant short-run asset effects, with no one-year effect exceeding .019. In contrast, the early studies by Albert Ando and Franco Modigliani indicated asset effects between .04 and .08, with no distinction between short-run and long-run effects, and while their most recent results incorporated in the *MPS* model point to considerably smaller short-term effects, at least for stock holdings, they are still substantially above the estimates reported in the other time-series analyses.[4]

---

[2] A description of these surveys and of the data collected appears in Dorothy Projector and Gertrude Weiss, and Projector.

[3] See, for example, Jean Crockett and Friend (1967), and Projector.

[4] The updated version of the *MPS* model represents revisions incorporated as late as January 1973. As compared with earlier estimates, which did not break down wealth into stocks and other assets, the long-run asset coefficient (assumed to be the same for both forms of assets) did not change appreciably. However, a much lower proportion of the effect on consumption of changes in the value of stock and a much higher proportion of the impact of changes in the value of other assets take place in the first three months.

II. Limitations of New Data

There are several difficulties in extending to the corresponding time-series relationships the results of an analysis of survey data on personal saving, income, capital gains, and assets derived from a single cross-section of households. First, since transitory components of some of these variables are more important in cross-section data, there is a need to separate normal from transitory components. Second, since we are interested in drawing intertemporal conclusions from interpersonal data, tastes effects must be considered. Third, measurement errors are generally larger for cross-section than for aggregate time-series data (see Friend and Stanley Schor). However, for the value of stock assets held by households and the associated capital gains, the cross-section data as provided by the *FRB* are likely to be fully as accurate as the aggregate time-series data in the United States—at least for the high proportion of respondents who reported the number of shares owned of each stock in their portfolios.[5] Moreover, as indicated subsequently, the blown-up *FRB* sample data on saving, income, stock assets, and capital gains are not too different from the corresponding national aggregates, once the sample is adjusted for deficient responses. Fourth, the definition of saving used by the *FRB* differs from the time-series (and customary economic) definition in several respects. The two most important differences are the absence of any deduction for depreciation on owned homes and of any allowance for saving in the form of life insurance. In our statistical results, we have consistently subtracted depreciation from the *FRB* measure of saving, and in part of our analysis we have added back a rough allowance for saving in the form of life insurance.[6] Fifth, while trust accounts are included in the household's net worth, capital gains on stock held in such accounts are not included in the household's capital gains, so that capital gains are somewhat understated and the capital gains effect somewhat overstated.[7]

Since the major contribution of this paper is to make use of the unique *FRB* data on the detailed stock holdings of individual households for deriving new estimates of capital gains effects on saving, it should be noted that the capital gains data, though of relatively high quality, are still subject to some error. Capital gains are estimated from the prices at the beginning and end of the year for individual stocks held by each household at the beginning of the year. Even where respondents correctly reported the number of shares owned of each stock in their portfolios, our estimates of accrued capital gains (or losses) attribute to a household gains which accrued to a stock after it had been sold, while accrued gains on purchases made during the year are excluded. Even so, the error introduced is rather small for individuals who provided data on their individual stock positions.[8]

---

[5] A fairly close correspondence between reported and actual ownership for respondents who give the number of shares owned is indicated in validation tests discussed in Robert Ferber et al. In contrast a recent revision of the estimated market value of outstanding stocks in the United States made as part of the Securities and Exchange Commission (*SEC*) *Institutional Investor Study* showed very little difference between the earlier *SEC* figures and the revised figures for the early 1960's, but a $283 billion or 37 percent upward revision by 1968, most of the discrepancy occurring after 1964.

[6] Depreciation on housing was estimated at 1.5 percent of the initial value of housing on the basis of the 1950 *FRB* "Survey of Consumer Finances." Where the survey definition of saving was expanded to include life insurance, the change in cash surrender value was used as a measure of saving in life insurance.

[7] It was possible to roughly estimate capital gains on the aggregate of stock held in trust accounts for each household reporting such accounts and to redo the analysis on this basis. Fortunately, the capital gains and other coefficients presented in Table 1 were virtually unaffected.

[8] As measured by the *FRB* survey, realized capital gains on securities (which includes some gains on assets

In some cases no identification of specific shareholdings was given, raising more serious questions of data reliability.[9] To guard against significant errors, for much of our analysis the sample used was restricted to those households which either reported no shareholdings or for which identified shareholdings accounted for at least 90 percent of the total value of their reported stockholdings. This restriction reduced the sample by 389 to 1,770 observations. For part of the analysis other restrictions were also employed reducing this sample further. When stock was unidentified, capital gains were estimated at 20 percent for the year.[10] As a result, the capital gain variable contains an additional small random error which would tend to bias somewhat its estimated coefficient towards zero. The sensitivity of the capital gain coefficient to random errors is tested in our statistical analysis.

Finally, the weighting procedures followed are important in the interpretation of the results, because the survey heavily oversampled upper income groups and also because nonrandom subsamples are used in much of our analysis. Roughly 128 weight classes distinguished by income, age, retirement status, and moving status (which indicates whether families moved in 1963) are employed, with each observation assigned a weight to represent a specific number of households in the population. Whenever observations are deleted, the remaining observations are proportionately scaled upward so that the total weight for each class remains unaffected. Should a weight class lose all of its observations, its weight is distributed among the members of a nearby weight class.

III. New Statistical Results

A number of cross-section regressions were estimated from the *FRB* survey data using the general form

$$S = a + bY_N + cY_T + dZ + eA_{-1} + f\,\text{Age} + g\ \text{Family Size} + h\ \text{Occupation}$$

where S is saving for 1963 net of depreciation on owned homes; Y_N, normal disposable income in 1962 and 1963;[11] $Y_T = Y_{63} - Y_N$; Z, 1963 accrued capital gains on stocks;[12] A_{-1} is market value of net worth at the end of 1962. Age and Family Size are continuous variables while Occupation is a dummy taking on the value 1 for households with self-employed heads and 0 for all other households.[13] In some regressions, V_{-1}, the net worth of stocks, and O_{-1}, the net worth of other assets, are substituted for A_{-1}.

The *FRB* survey did collect data on realized capital gains on all securities combined as a global figure for each household, but these figures seemed extremely weak and could not be checked against other data so that in the results presented in this paper no distinction was made between realized and unrealized capital gains.[14] No

---

other than stocks) represent only about 5 percent of the value of accrued capital gains on stocks. Furthermore, gains attributed to common stock which was sold will be offset to a considerable extent by the gains which accrued to those individuals who reinvested the proceeds of their sales.

[9] Ferber et al. found that in these cases the wealthy underreported their holdings whereas the (relatively) poor overstated their holdings.

[10] The composite 1963 New York Stock Exchange (*NYSE*) capital gain was 18.1 percent and the over the counter (*OTC*) gain was 20.4 percent as determined respectively from the New York Stock Exchange Common Stock and National Quotation Board indices.

[11] Normal income was also defined as the expected value of disposable income derived by regressing 1963 household income on the average incomes of each of a number of socio-economic-demographic groups to which the household belonged, with no important changes in the short-run asset effects presented here.

[12] Accrued gains in 1963 are gains reflecting the price changes during that year.

[13] Dummy variables were also introduced for age, distinguishing among household heads who were less than 39 years old, 40 to 59 years, and 60 and over years, without any noteworthy change in regression results.

[14] External sources as well as the survey data indicate that realized capital gains on stock are a small fraction

variable was introduced for accrued capital gains on assets other than stocks since such data were not available from the *FRB* survey. As a result, we were not able to check the intuitively implausible *MPS* result that the effect of an unexpected as well as expected change in value of homes, bonds, and other nonequity assets on household consumption takes place immediately (i.e., in the same quarter). We were, however, able to check the implications of the *MPS* model for the timing pattern of consumption effects flowing from changes in the market value of stock owned.

Theory suggests that the capital gains coefficient d should be between 0 and the negative asset coefficient e.[15] For unanticipated capital gains, the coefficient would be expected to be close to zero; for anticipated gains, close to e. However, even though investors are likely to have some long-run capital gains rate in mind when they invest, short-run capital gains may be almost entirely unanticipated. Since any anticipated capital gains are presumably highly correlated with initial assets (and theoretically would be expected to have the same coefficient), the d coefficient in the saving regressions should reflect mainly the effect of unanticipated gains. We are interested, of course, in both the d and e coefficients, measuring respectively the short-run (same year's unanticipated capital gain) and long-run asset effects on saving.

Unfortunately, there is reason to believe that only the capital gains coefficient is measured with any precision. The assets or net worth coefficient is likely to be substantially downward biased (in absolute value) by tastes effects, since households which have a high propensity to save will have both high savings and high assets. However, the capital gains coefficient is likely to be much less biased by tastes effects since tastes are much more highly correlated with assets than with capital gains. Indeed, when assets as well as capital gains are included in the savings regression, the correlation of capital gains and tastes should be close to zero.

Table 1 presents the major statistical results obtained from the analysis of the *FRB* survey, with data for each household weighted by the appropriate sampling ratio. The definitions of the saving, income, capital gains, and asset variables used in these regressions are those specified by the *FRB* (see Projector) except that depreciation on owned homes was deducted both from saving and assets. While this depreciation adjustment was made to improve conceptually the saving and wealth estimates, it should be noted that the capital gains coefficient was not affected appreciably. The most important remaining deficiency in the measures of saving and assets generally used in Table 1 is that the change in and the level of equity in insurance are not reflected in saving or assets. However, several regressions were computed in which the change in and the level of the cash surrender value of life insurance were added to the saving and asset variables, and one of these regressions is presented in the table. Equation (1) in the table, which covers all 2,159 households in the sample, points to a low depressant effect of capital gains Z and assets A_{-1} on saving, a rather high normal income Y_N effect, and a low transitory income Y_T effect. Only the Y_N and A_{-1} coefficients are statistically significant at the customary 95 percent level, and the goodness of fit ($\overline{R}^2$) for the equation as a whole is quite low, even for ungrouped cross-section data.

Since there are, except for the respondents who give the names and quantities

---

of unrealized gains. See Lieberman for the effects of realized capital gains.

[15] For a theoretical discussion of the signs and values of these parameters, see Friend and Paul Taubman, and Crockett and Friend (forthcoming).

TABLE 1—HOUSEHOLD SAVINGS REGRESSIONS FOR DIFFERENT SAMPLES

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Size of Sample	2,159	1,770	1,770	1,195	1,195	1,195	303	303	1,195	1,195	1,195	1,195
Constant	−0.21	−0.50	−0.21	−0.30	−0.38	−0.15	0.73	−0.82	−0.40	−0.54	−0.41	−0.10
	(−0.27)	(−0.59)	(−0.25)	(−1.01)	(−1.27)	(−0.45)	(0.77)	(−0.97)	(−3.68)	(−2.93)	(−2.28)	(−0.21)
Y_N	0.24	0.23	0.19	0.10	0.11	0.11	0.04	0.09	0.11	0.18	0.11	0.11
	(5.84)	(5.01)	(4.03)	(6.06)	(6.70)	(6.17)	(1.16)	(2.64)	(7.86)	(10.27)	(6.28)	(4.80)
Y_T	0.14	0.41	0.47	0.32	0.30	0.28	0.33	0.35	0.24	0.32	0.29	0.51
	(1.08)	(2.73)	(3.16)	(6.60)	(6.29)	(5.38)	(12.10)	(3.93)	(5.57)	(4.66)	(6.42)	(7.19)
Z	−0.018	−0.040	−0.040	−0.028	−0.041	−0.024	−0.015	−0.021	−0.019	−0.029	−0.019	−0.029
	(−1.40)	(−3.77)	(−3.76)	(−4.85)	(−6.31)	(−3.95)	(−1.30)	(−2.19)	(−1.05)	(−4.88)	(−2.10)	(−8.70)
V_{-1}			−0.021		0.005							
			(−11.18)		(3.84)							
O_{-1}			0.000		−0.005							
			(0.05		(−2.83)							
A_{-1}	−0.014	−0.016		0.001		0.001	0.001	0.001	−0.002	0.000	−0.001	0.002
	(−8.23)	(−9.80)		(1.20)		(0.20)	(0.48)	(0.58)	(−0.64)	(0.32)	(−0.72)	(3.26)
Age	−0.002	0.002	−0.004	0.001	0.003	−0.002	−0.010	0.009	0.002	0.001	0.002	−0.001
	(−0.20)	(0.18)	(−0.29)	(0.25)	(0.71)	(−0.38)	(−0.79)	(0.81)	(1.25)	(0.54)	(0.83)	(−0.19)
Family Size	−0.119	−0.100	−0.085	−0.016	−0.021	−0.07	−0.004	0.067	−0.008	−0.055	−0.011	−0.026
	(−1.30)	(−1.02)	(−0.87)	(−0.43)	(−0.54)	(−0.16)	(−0.03)	(0.60)	(−0.56)	(−2.05)	(−0.55)	(−0.58)
Occupation	1.08	0.85	0.33	0.47	0.62	0.70	1.42	1.44	0.30	0.16	0.31	0.14
	(1.53)	(1.09)	(0.42)	(1.81)	(2.38)	(2.52)	(2.39)	(2.31)	(2.69)	(1.02)	(2.32)	(0.59)
$\bar{R}^2$	.035	.061	.075	.099	.112	.088	.422	.106	.094	.134	.087	.141
Mean of Dependent Variable ($000)	0.472	0.388	0.388	0.389	0.389	0.493	−0.032	0.422	0.214	0.432	0.216	0.526
Standard Error ($000)	0.165	0.179	0.178	0.063	0.062	0.067	0.141	0.134	0.027	0.050	0.033	0.078

Note: Dependent variable is saving for 1963 net of depreciation on owned homes and in equation (6) is adjusted for saving in the form of life insurance; Y_N, normal income defined as the average of annual disposable income in 1962 and 1963; $Y_T=Y_a-Y_N$; Z, 1963 accrued capital gains on stocks; V_{-1}, the market value of corporate equities at the end of 1962; O_{-1}, the market value of assets other than equities at the end of 1962; A_{-1}, market value of assets or net worth held at the end of 1962; and Age and Family Size are continuous variables while Occupation is represented by a dummy variable taking on the value 1 for households with self-employed heads and 0 for all other households. Figures in parentheses are t-statistics. Dollar variables are measured in thousands of dollars. $\bar{R}^2$ is R^2 adjusted for degrees of freedom. Standard error in ast line is standard error of expected saving conditional on mean values of independent variables.

of individual shares owned, substantial errors in reporting the value of stock assets and in estimating capital gains on stock,[16] the two variables in which we are particularly interested, a second set of regressions, equations (2)–(3), was based on those households which gave satisfactory detail on stock ownership. This second sample consisting of 1,770 households excluded those respondents reporting stock ownership who did not stipulate the amounts of individual stocks held accounting for at least 90 percent of the value of all stock reported owned by the household. Equation (2) is directly comparable to (1), except for the smaller sample, and (3) substitutes the market value of stocks V_{-1} and other assets O_{-1} for total initial assets.

The results of equations (2) and (3) represent an improvement over equation (1), with a moderately higher overall correlation, and with more generally satisfactory regression coefficients. Thus the coefficients of Y_T are now statistically significant and, as would be expected, are higher than those of Y_N. Of special importance for our analysis, the coefficients of Z not only have the theoretically anticipated sign but are significant as well. Of the other economically important variables, the coefficient of A_{-1} is increased somewhat (in absolute value). The substitution of V_{-1} and O_{-1} for A_{-1} gives unreasonable results for the coefficient of O_{-1}, perhaps partly as a reflection of the multicollinearity between V_{-1} and O_{-1} and partly as a reflection of very large random errors in the measurement of O_{-1}. (Of the other socio-economic-demographic variables, none is significant though there is some indication that small families and entrepreneurs save more than other households.)

---

[16] See Ferber et al. Even for respondents giving the number of shares of all stock reported owned, there is some tendency to overstate the size of very small holdings and understate very large holdings, and thus to bias somewhat upward (in absolute value) the coefficient of Z.

To eliminate other apparent sources of error in measuring saving, income, and assets, households which did not pass a number of quality checks were excluded resulting in a reduced sample of 1,195 households, and a new set of regression results is presented in equations (4)–(6). Households were excluded if 1) net worth was negative, 2) normal income was less than $1,000, 3) the absolute level of saving was greater than normal income, and 4) stock assets at the beginning of year had a market value of less than $200,000 and capital gains were in excess of the value of stocks held, or stock assets had a market value of $200,000 or more and capital gains were in excess of half the value of stock held. The first of these exclusions was the most important and the fourth quite unimportant in reducing the sample size. The third is by far the most important exclusion in terms of impact on the regression coefficients.

Equations (4) and (5) for the 1,195 sample are the counterparts of (2) and (3) for the 1,770 sample. It will be noticed that the normal income and asset coefficients are substantially reduced, and the asset coefficient is no longer significant,[17] but that the capital gains coefficients are much less affected. (The occupational dummy variable is now significant or close to significant, with entrepreneurs again saving more than other households.) Equation (6) is the same as (4) except that savings and assets are now expanded conceptually to include the change in and level of the cash surrender value of private life insurance.[18] This provides an approximate measure of household saving which conforms more closely to the measure used in the aggregate national income statistics and econometric models. The adjustment of saving and assets for private life insurance does not significantly affect the regression coefficients either here or in other equations tested, though the capital gains coefficient is somewhat lowered.

To eliminate one last potentially significant source of error in measuring assets and capital gains, households which reported owning no stock were excluded from the preceding sample of 1,195 households, resulting in a much smaller sample of 303 households. The reason for this exclusion is that validation checks suggest that a number of households reporting no stock do in fact own some (see Ferber et al.). On the other hand, though the quality of the sample is further improved, its size and representativeness are greatly reduced and the resulting sample is biased in favor of households which have a propensity for saving in the form of stock. While the mean saving, income, net worth, and capital gains of households are affected only moderately as the sample is reduced in size to 1,195 households, the mean market value of stock held and capital gains (though not saving and income) are increased substantially as the sample is further reduced to 303 households.

In equation (7), the 303 sample households are weighted using the 128 weight classes which result from the *FRB* survey weighting procedures. In view of the small size of the remaining sample relative to the number of weight classes, obviously few or no observations are included in many of the classes if the *FRB* weighting procedures are followed. To handle this problem, the number of weight classes was reduced to 12 based upon normal income. The Y_N weights are internally generated by adding up the *FRB* weights of all the individual households within the larger normal income classes. Class sizes were chosen so as to contain an approxi-

[17] In one equation, the stock assets coefficient V_{-1} is significant but has the wrong sign.

[18] It would be preferable to use equity instead of cash surrender value, but the data are not available in the *FRB* survey.

mately equal number of observations. It is noteworthy that when the 303 sample is weighted by normal income, the blown-up survey estimates of mean saving and income are reasonably close to the external figures obtained from time-series data, which is also true for the larger samples. However, the capital gains and market value of stock estimates from the 303 sample are appreciably larger than the corresponding external figures,[19] whereas the estimates from the larger samples are substantially smaller than the external figures. As a result, the capital gains effects indicated by the 1,195 and larger samples might be expected to be somewhat overstated while the reverse would be true for the 303 sample.

Fortunately, when the Y_N weights were used in equation (8), the statistical results were not too different from those obtained in the corresponding equation (4) based on the 1,195 sample. The most obvious difference between equations (7) and (8) or (4) is in the correlation, with $\bar{R}^2$ for (7) much higher than for the others. However, little confidence can be placed in (7) in view of the large weights ascribed to a small number of statistically biased observations in many of the 128 weight classes and the highly questionable figure for mean saving.

Finally, since the first eight equations in Table 1 all display significant heteroscedasticity in the residuals so that the t-values are overstated, four other equations derived from the 1,195 sample have been added (equations (9)–(12)) which correct for this problem by dividing S, Y_N, Y_T, Z, A_{-1}, and the constant term either by $A_{-1}^{.25}$ (equations (9) and (10)) or by $Y_N^{.25}$ (equations (11) and (12)).[20] Equations (9) and (11) like the first eight equations represent weighted regressions, while (10) and (12) are unweighted. The first three of these four equations almost completely eliminate heteroscedasticity in the residuals while the last, equation (12), substantially reduces heteroscedasticity but it is still statistically significant. The estimated capital gains effect on saving in these four regressions is somewhat lower on average than in the corresponding equation (4) which is weighted but not adjusted for heteroscedasticity, but the differences are probably not statistically significant.

Since age is assigned an important role in saving by life-cycle theories, equation (4) which is based on the 1,195 sample was recomputed for each of five age groups and the results presented in Table 2. The capital gains coefficients show substantial variability, but are never significant. Hence, they are not as satisfactory as the more aggregate results. However, all the capital gains coefficients have the correct sign and the variability may well reflect the small size of the subsamples.[21]

The estimation of other mathematical forms did not change the range of Z coefficients from those indicated above. As suggested earlier, measurement error would be expected to introduce some downward bias in the coefficient of Z, but this bias is likely to be small at least in the 1,195 and 303 samples.[22]

---

[19] Mean saving (exclusive of insurance), family disposable income, and capital gains on stock for the 303 sample were $422, $5,720, and $1,911, respectively, per household. Multiplying these figures by 57.9 million households (Projector, p. 320) provides 1963 aggregate personal saving and family disposable income fairly close to the corresponding national accounts totals, but aggregate capital gains on stock appreciably higher than the product of the *SEC* total of individuals' stock holdings and the 20 percent estimated percentage capital gain.

[20] This is equivalent to assuming that for different values of these independent variables the ratio of the original variance of residuals to $A_{-1}^{.5}$ or $Y_N^{.5}$ is constant.

[21] When the sample households are classified into three instead of five age groups, the coefficient of capital gains for the middle age group, viz., 40–59 years for the age of household head, is statistically significant (with a coefficient of $-.039$ and t-value of 4.24). The omission of the age variable from either the five or three age group regressions has virtually no effect on the Z coefficients, nor does the substitution of dummy age variables (intercept and slope) for age as a continuous variable.

[22] The sensitivity of the Z coefficient to measurement error was tested by adding to Z a random component

TABLE 2—HOUSEHOLD SAVINGS REGRESSION BY AGE OF HEAD

Age of Household Head	Constant	Y_N	Y_T	Z	A_{-1}	Age	Family Size	Occupation	$\bar{R}^2$
−34	−2.10	.20	.38	−.165	.018	.048	−.034	.04	.236
	(−2.70)	(4.57)	(3.65)	(−.26)	(1.81)	(1.57)	(−.50)	(.08)	
35–44	3.78	.04	.23	−.034	−.001	−.092	−.035	1.59	.098
	(1.68)	(1.07)	(2.18)	(−1.62)	(−.16)	(−1.76)	(−.39)	(3.00)	
45–54	.63	.12	.33	−.039	−.003	−.009	−.143	.71	.049
	(.22)	(3.12)	(2.49)	(−.79)	(−1.19)	(−.15)	(−1.50)	(1.28)	
55–64	−.18	.11	.20	−.003	.004	−.001	−.014	−.90	.173
	(−.05)	(3.02)	(2.29)	(−.18)	(3.36)	(−.02)	(−.13)	(−1.44)	
65+	.95	.11	1.15	−.010	−.003	−.013	−.065	−.84	.161
	(.53)	(2.47)	(6.00)	(−.89)	(−1.09)	(−.56)	(−.56)	(−.99)	

Note: See Table 1 for explanation of symbols.

Although a comparison of three of our coefficient estimates—Z, V_{-1}, and O_{-1}—may be made with the *MPS* model, those of V_{-1} and O_{-1} are likely to be substantially biased downward (in absolute value) as a result of the impact of tastes, as noted earlier. For this reason, the fact that the coefficients of V_{-1} and O_{-1} in the *MPS* model are substantially larger than those provided by the analysis of survey results is not significant evidence that the *MPS* results are biased upward.

In contrast, the coefficients of Z in the analysis of 1963 survey results are likely to be little affected by tastes effects. The Z coefficients estimated from the survey regressions, ranging from −.015 to −.041, may be compared with corresponding estimates of −.040 for the last published version of the *MPS* model, −.032 from the latest unpublished version, and much lower figures from other time-series regressions.[23] Thus, the *FRB* survey results suggest that the latest *MPS* estimate of the first year's effect of a change in the value of stock assets on saving and consumption is not far off from the true figure though it may be a little on the high side, while the corresponding effects implied by most other time-series results seem too low.

---

normally distributed with mean zero and standard deviation $k\sigma_t$ where k varied between .1 and .5 and σ_t, the standard deviation of capital gains, is $13,701. Five simulations of equation (4) of Table 1 were run for each value of k. For k equal to .1 the mean value of the Z coefficient was −.028 with a standard deviation of .00049. The corresponding figures were −.022 and .00148 for $k=.4$, and −.020 and .00163 for $k=.5$. At least for the 1,195 and 303 sample sizes, we believe the data to be measured with sufficient accuracy that k is likely to be well under .5.

[23] The estimate for the impact of unanticipated capital gains on saving in the *MPS* model was obtained by assuming that the *MPS* quarterly capital gains coefficients are equal to a weighted combination of the fully anticipated *MPS* wealth effect (.0544 in the latest version of the model) and the unknown unanticipated capital gains effects. The effect in subsequent periods of unanticipated capital gain in any quarter is the product of the weight of unanticipated gains in that quarter multiplied by the appropriate quarterly coefficient of such gains in the sequence of coefficients reflecting adjustment of households to the new level of assets. This formulation leads to five unknowns if the weight of unanticipated capital gains is assumed constant (the four quarterly unanticipated capital gains coefficients and the weight of unanticipated—or anticipated—capital gains) but only four equations (one for each *MPS* quarterly coefficient). If the effect of contemporaneous capital gains is zero, an upper bound of .3 is found for the weight of anticipated capital gains in the latest version of the *MPS* model. Assuming values between 0 and .3 for the weight of anticipated capital gains results in a narrow bracketed solution for the quarterly impact of unanticipated capital gains permitting use of the midpoints. The midpoints of the bracket for each quarter may then be applied to the percentage of 1963 capital gains which accrued during each quarter to estimate the impact of unanticipated capital gains on 1963 savings. A simple average of the *NYSE* and *OTC* composite index was used to calculate the percentage of 1963 capital gains which accrued during each quarter.

However, it should be noted that the time-series estimates represent an average over a period of years while the survey results presented here are for a single year only.

To obtain a more definitive result for the first year's effect in general, survey data for other periods would be needed, and to obtain a complete answer to the long-term as well as short-term time sequence of asset effects would require panel survey data. The statistical findings in this paper not only strongly support the inclusion of capital gains, and implicitly net worth, in the consumption function, but also indicate that survey data constitute an unusually promising source of information for quantifying the time-sequence of asset effects on saving and consumption.

REFERENCES

A. Ando and F. Modigliani, "The 'Life Cycle' Hypothesis of Saving: Aggregate Implications and Tests," *Amer. Econ. Rev.*, Mar. 1963, *53*, 55–84.

——— and ———, "The 'Life Cycle' Hypothesis of Saving: A Correction," *Amer. Econ. Rev.*, Mar. 1964, *54*, 111–13.

J. J. Arena, "Capital Gains and the 'Life Cycle' Hypothesis of Saving," *Amer. Econ. Rev.*, Mar. 1964, *54*, 107–11.

———, "Postwar Stock Market Changes and Consumer Spending," *Rev. Econ. Statist.*, Nov. 1965, *47*, 379–91.

K. B. Bhatia, "Capital Gains and the Aggregate Consumption Function," *Amer. Econ. Rev.*, Dec. 1972, *62*, 866–79.

J. Crockett and I. Friend, "Consumer Investment Behavior," in R. Ferber, ed., *Determinants of Investment Behavior*, Universities–Nat. Bur. Econ. Res. conference series, New York 1967, 15–127.

——— and ———, "Consumption and Saving in Economic Development," in *Proceedings of International Conference on Consumption, Income and Prices*, Washington forthcoming.

M. K. Evans, "The Importance of Wealth in the Consumption Function," *J. Polit. Econ.*, Aug. 1967, *75*, 335–51.

M. S. Feldstein, "Tax Incentives, Corporate Saving and Capital Accumulation in the United States," *J. Publ. Econ.*, Apr. 1973, *2*, 159–71.

——— and G. Fane, "Taxes, Corporate Dividend Policy and Personal Savings: The British Postwar Experience," *Rev. Econ. Statist.*, Nov. 1973, *55*, 399–411.

R. Ferber et al., "Validation of Consumer Financial Characteristics: Common Stock," *J. Amer. Statist. Assn.*, June 1969, *64*, 415–32.

I. Friend and S. Schor, "Who Saves?," *Rev. Econ. Statist.*, May 1959, *41*, 213–48.

——— and P. Taubman, "The Aggregate Propensity to Save: Some Concepts and Their Application to International Data," *Rev. Econ. Statist.*, May 1966, *48*, 113–23.

C. Lieberman, "An Investigation of the Short-Run Effects of Capital Gains on Household Consumption and Saving," unpublished doctoral dissertation, Univ. Pennsylvania 1974.

F. Modigliani, "Monetary Policy and Consumption," in *Consumer Spending and Monetary Policy: The Linkages*, Fed. Reserve Bank Boston, *Monetary Conference Series*, no. 5, Boston, June 1971, 9–85.

D. S. Projector, *Survey of Changes in Family Finances*, Washington 1968.

——— and G. S. Weiss, *Survey of Financial Characteristics of Consumers*, Washington 1966.

Board of Governors of the Federal Reserve System, *Fed. Reserve Bull.*, "Survey of Consumer Finances," Nov. 1960, *46*, 1454.

MIT-PENN-SSRC Econometric Model of the United States, unpublished mimeo, January 1973, Wharton Econometric Forecasting Associates.

National Quotation Bureau, OTC Industrial Stock Average, 1962–63.

New York Stock Exchange, Composite Stock Index, 1962–63.

Securities and Exchange Commission, *Institutional Investor Study*, supp. Vol. 1, Hearings, 92d Congress, 1st Session, House Document No. 92–64, Part 6, Washington 1971.

U.S. Office of Business Economics, *Surv. Curr. Bus.*, July 1966, *29*, Table 57.

Pure Traded Intermediate Products and The Heckscher-Ohlin Theorem

By ALBERT G. SCHWEINBERGER*

In a recent paper, Raveendra Batra and Francisco Casas introduced a nontraded intermediate product into the basic Heckscher-Ohlin trade model. This is an important step forward, because nontraded intermediate products (as well as intermediate products in general) must be of great quantitative significance in the real world. Traditionally, intermediate products have for a long time been ignored in trade theory. When traded intermediate products were finally taken account of, it seemed that most of trade theory could still be applied with only rather minor changes. This initial "optimism" has now given way to a much more cautious attitude. It has been realized that one of the main reasons why it seemed at first sight that most results of trade theory are still valid when intermediate products are allowed for is to be found in the rather special way in which intermediate products were introduced:

1) Intermediate input coefficients were assumed to be fixed.
2) The intermediate products were of the Leontief kind, i.e., they were products which are used for production as well as for final consumption purposes.

The paper by Batra and Casas is exceptional in that it introduces pure intermediate products into the Heckscher-Ohlin framework. One of the key results of their paper is that the Heckscher-Ohlin theory remains valid in this new model only if factor intensities are interpreted in terms of total original input requirements. If only direct original factor input requirements are taken into account, the Heckscher-Ohlin model may break down. The present paper attempts to throw some light on the validity of the Heckscher-Ohlin theorem when the model contains pure *traded* intermediate products. Furthermore, the assumption of fixed intermediate input coefficients, which is so prominent in most trade theory (for example, effective protection theory), is also relaxed.

The paper is divided into two sections. In Section I, a model is analyzed where the number of produced products (for the world as a whole) is greater than the number of scarce primary factors. In Section II, both countries are assumed to produce one final commodity and one pure intermediate product in the original pretrade equilibrium and under conditions of universal free trade.

I. The Model

At first sight it would seem that the introduction of pure traded intermediate products into a traditional two original factor Heckscher-Ohlin trade model, where both final commodities are produced under constant returns to scale, is quite impossible. The reason for this is that in a traditional Heckscher-Ohlin model the trading pattern is determined by the pretrade relative prices and it is assumed that both countries produce all commodities in the

---

* University of East Anglia. I am indebted to an unknown referee for helpful comments and criticism. An early version of this paper was presented to the International Economics Study Group at the London School of Economics.

pretrade and in the free trade equilibrium. But with three traded commodities being produced in each country, it follows that the traditional assumptions of the Heckscher-Ohlin theorem imply that under free trade factor prices are equalized. In what follows it has been assumed that free trade does not result in an equalization of factor prices. Given the assumption of identical production functions and constant returns to scale, it would therefore seem that allowing for two final commodities, two original factors of production, and one traded intermediate product, the following set of functions cannot be independent:

$$\begin{aligned} a_{11}p_1 + a_{21}p_2 + a_{m1}p_I &= q_1 \\ a_{12}p_1 + a_{22}p_2 + a_{m2}p_I &= q_2 \\ a_{13}p_1 + a_{23}p_2 \qquad\quad &= p_I \end{aligned}$$

where a_{ij} = primary factor input coefficients
a_{mj} = intermediate input coefficients
p = primary factor prices
q = free trade equilibrium final output prices
p_I = the free trade equilibrium price of the intermediate product

Let q and p_I be given. It then follows that the set of three cost functions cannot be independent as long as cost functions, and therefore production functions, are homogeneous of degree 1. At this point one may be tempted to give up. But there is a way out. While it follows from the above that both countries cannot normally produce all commodities in a trading equilibrium, it does not follow that the world as a whole may not be producing all three commodities. If primary factor endowments are different in the two countries, there is no reason at all why there should not be specialization between countries in the trading equilibrium. It seems likely that the assumption of nonspecialization between countries, which is so prevalent in much of the literature on trade theory, is one of its main drawbacks for the simple reason that it is specialization rather than nonspecialization which one appears to observe in the real world. Probably the main reason why nonspecialization has almost always been assumed is one of analytical convenience. Since the rise of concave programming, a technique of analysis has become available which I hope to show can easily cope with the problems of particular specialization patterns.

It is therefore suggested that the assumption of nonspecialization of the traditional Heckscher-Ohlin model be dropped, but the assumption of identical production functions retained. It follows from this that in general factor prices will not be equalized between countries, because if the specialization pattern of production is different between the two countries, one is solving two different sets of cost functions. It has long been observed in practice that primary factor prices are not equalized by trade. The traditional conclusions have been that either production functions are not identical or factor intensity reversals occur over the relevant range of factor prices. Perhaps a much more important reason for the nonequalization of primary factor prices by free trade is to be found in the presence of particular specialization patterns between countries.

In order to allow for a particular specialization pattern to emerge in free trade equilibrium, one has to drop the traditional approach of cost functions which are all fulfilled and rewrite the problem as a non-linear program. Factor prices in the two countries will be determined as the optimal solution vectors of the two following non-linear programs:

Country 1: Min $p_1K + p_2L$

subject to

$$\begin{aligned} a_{11}p_1 + a_{21}p_2 + a_{m1}p_I &\geq q_1 \\ a_{12}p_1 + a_{22}p_2 + a_{m2}p_I &\geq q_2 \\ a_{13}p_1 + a_{23}p_2 \qquad\quad &\geq p_I \\ p_1,\ p_2 &\geq 0 \end{aligned}$$

where K and L are primary factor endowments in country 1.

$$\text{Country 2:}\quad \text{Min } p_1K' + p_2L'$$

$$\begin{aligned} \text{subject to}\qquad & a_{11}p_1 + a_{21}p_2 + a_{m1}p_I \geq q_1 \\ & a_{12}p_1 + a_{22}p_2 + a_{m2}p_I \geq q_2 \\ & a_{13}p_1 + a_{23}p_2 \qquad\qquad \geq p_I \end{aligned}$$

where K' and L' are primary factor endowments in country 2.

Perhaps it should be made explicit that under constant returns to scale both these programs should in general be soluble, because the variables of all three constraints are primary factor prices only. From inspection of the two programs, it follows that the world as a whole will produce all three commodities with primary factor prices not being equalized by free trade, if and only if there is a minimum difference between the factor endowment vectors of the two countries (given that production functions are identical). If factor endowment vectors do not differ by this minimum amount (this will be clarified later), particular specialization patterns between countries could not emerge unless production functions are different between countries. In the present analysis the assumption of identical production functions is retained. There are three more assumptions which have to be made in order to ensure that in this model primary factor prices are uniquely determined for given primary factor endowments and prices q_1, q_2, and p_I:

1: All production and therefore cost functions are homogeneous of degree 1.
2: All unreduced cost functions are concave.
3: Each country produces two products.

Assumption 2 implies that:

$$(1)\quad \frac{\partial a_{13}}{\partial p_1} \leq 0 \qquad \begin{vmatrix} \dfrac{\partial a_{13}}{\partial p_1} & \dfrac{\partial a_{13}}{\partial p_2} \\ \dfrac{\partial a_{23}}{\partial p_1} & \dfrac{\partial a_{23}}{\partial p_2} \end{vmatrix} \geq 0$$

$$(2)\quad \frac{\partial a_{1i}}{\partial p_1} \leq 0 \qquad \begin{vmatrix} \dfrac{\partial a_{1i}}{\partial p_1} & \dfrac{\partial a_{1i}}{\partial p_2} \\ \dfrac{\partial a_{2i}}{\partial p_1} & \dfrac{\partial a_{2i}}{\partial p_2} \end{vmatrix} \geq 0$$

$$\begin{vmatrix} \dfrac{\partial a_{1i}}{\partial p_1} & \dfrac{\partial a_{1i}}{\partial p_2} & \dfrac{\partial a_{1i}}{\partial p_I} \\ \dfrac{\partial a_{2i}}{\partial p_1} & \dfrac{\partial a_{2i}}{\partial p_2} & \dfrac{\partial a_{2i}}{\partial p_I} \\ \dfrac{\partial a_{mi}}{\partial p_1} & \dfrac{\partial a_{mi}}{\partial p_2} & \dfrac{\partial a_{mi}}{\partial p_I} \end{vmatrix} \leq 0$$

where $i = 1, 2$.

We now want to prove that the conditions stated immediately above imply that the cost functions of assumption 2

$$q_i = q_i(p_1, p_2, p_I) \qquad i = 1, 2$$

are concave in the variables p_1 and p_2 if p_I is fixed at some level p_I^0. In fact, the proof is very easy. Note that a cost function is concave if, for given traded final output and intermediate product prices, $d^2p_2/dp_1^2 \geq 0$. One immediately obtains (assuming cost minimization):

$$\frac{dp_2}{dp_1} = -\frac{a_{11}}{a_{21}}$$

so that

$$\frac{d^2p_2}{dp_1^2} \geq 0, \qquad \text{for } p_I = p_I^0$$

follows immediately from the conditions stated under assumption 2.

Proceeding similarly with the cost functions of final commodity 2 and the intermediate product, the solution of the nonlinear program (for given q_1, q_2, and p_I) may be illustrated diagrammatically as in Figure 1.[1] After plotting the cost functions, the first step is to derive the feasible set of solution values p_1 and p_2. Making the assumptions 1, 2, and 3 stated above, it can readily be shown that the feasible constraint set must be convex (in Figure 1 it is the shaded area). Clearly, primary factor prices will be uniquely determined for given original factor endowments, provided that the constraint set is convex and that each country produces two products. It should be emphasized at this point that the cost functions, as shown in Figure 1, represent a rather special case. As can be seen, the boundary of the feasible set of original factor price solutions shows two vertices (A and B). The right- and left-hand derivatives dp_2/dp_1 at A and B indicate the limits within which variations of primary factor endowment ratios between the two countries are compatible with nonspecialization in production and an equalization of primary factor prices (assuming that free trade equilibrium prices q_1, q_2, and p_I remain unchanged).

---

[1] It should be carefully noted that the figure has been drawn on the assumption that one country is in equilibrium at the vertex A and the other at B, so that factor prices are not equalized by trade. If factor prices were equalized by trade, both countries could normally produce all three commodities. The necessary and suffi-

To discuss the validity of the Heckscher-Ohlin theorem when the assumption of nonspecialization is relaxed, the attention of the reader is drawn to Figure 1. Before embarking upon a more detailed analysis, it should be stated that the K/L ratios for final commodities 1 and 2 and the intermediate product for various factor price ratios (p_2/p_1) can quite easily be read off Figure 1. It may be shown, for example, that the slope of the q_1 curve at any point is equal to the direct K/L ratio used in the production of final commodity 1:

$$dq_1 = a_{m1}dp_I + a_{11}dp_1 + a_{21}dp_2$$

but $dq_1 = 0$ and $dp_I = 0$. Therefore, it follows that $dp_2/dp_1 = -a_{11}/a_{21}$, and similarly with the cost function of the other final

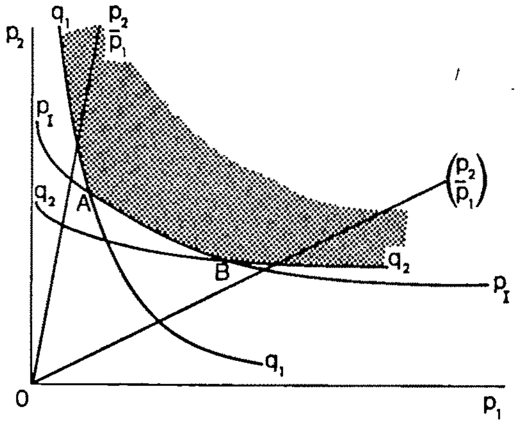


FIGURE 1

commodity and the traded intermediate product. We are concerned here only with the validity of the Heckscher-Ohlin theorem when factor endowments and demand functions are such that one country is in equilibrium at A and the other at B. It should also be noted that Figure 1 has been drawn in such a way that within the relevant primary factor price ratios (p_2/p_1) and p_2/p_1, there is a clearly defined ranking of direct primary factor intensities, such that $(K/L)_1 > (K/L)_I > (K/L)_2$. (It should perhaps be added here that whenever a pair of cost functions has at most one point of intersection within the relevant range of primary factor prices, no factor intensity reversals can occur. Also, if the ranking is as shown on Figure 1, a ranking of final commodities according to direct factor intensities must always be the same as a ranking according to total factor intensities. For a proof, see Batra and Casas.)

Let country 1, which is the relative capital abundant economy, be in equilibrium at A and let the other country be in equilibrium at B. There is now obviously no problem with the Heckscher-Ohlin theorem correctly predicting the trading pattern of final commodities. The country with the greater capital endowment will obviously export the more capital intensive final commodity. But both

---

cient conditions for the latter to be the case are: the endowment ratios of both countries must lie between any given pair of direct capital/labor ratios used in the production of the same two products; and both endowment ratios must also lie between the *total* capital/labor ratios of the two final commodities.

countries produce the intermediate product and it has to be proved that country 1 will export it, in order to establish the validity of the Heckscher-Ohlin theorem. For this purpose, in trading equilibrium, world supply must be equal to world demand for every commodity. World demand for the intermediate product is given by $a_{m1}C_1 + a'_{m2}C'_2$ and world supply by C_I and C'_I (where the primes denote country 2). If the intermediate product is to be imported by country 2, one must have

$$a_{m1}C_1 - C_I < 0$$

The factor market equilibrium conditions in country 1 are:

$$a_{11}C_1 + a_{1I}C_I = K$$

$$a_{21}C_1 + a_{2I}C_I = L$$

Solving for C_1 and C_I and substituting into $a_{m1}C_1 - C_I$, one obtains:

$$a_{m1}C_1 - C_I = \frac{\begin{vmatrix} K & a_{m1}a_{1I} + a_{11} \\ L & a_{m1}a_{2I} + a_{21} \end{vmatrix}}{|a|} < 0$$

As pointed out above, if the direct factor intensity ratio of the intermediate product is nonextremal, i.e., lies between the direct factor intensity ratios of the two final commodities, there is no problem with regard to the interpretation of the Heckscher-Ohlin theorem referring either to direct or total factor intensities of final commodities. In the traditional Heckscher-Ohlin model it is assumed that the capital/labor ratios of both countries lie within the convex cones of the capital/labor ratios used in the production of both commodities. If it is assumed that the same holds in the present model, i.e., total capital/labor ratio used in the production of commodity 1 > endowment ratios of both countries > total capital/labor ratio used in the production of commodity 2, it immediately follows that $a_{m1}C_1 - C_I < 0$, or, verbally, that country 1 (the capital abundant country) will export the intermediate product and country 2 (the labor abundant country) will import it. It must be emphasized that this result holds only if the endowment ratios of the two countries lie between the total original factor intensity ratios of final commodities. Clearly, it does not follow from $a_{11}/a_{21} > a_{1I}/a_{2I} > a_{12}/a_{22}$ that $K/L < (a_{11} + a_{m1}a_{1I})/(a_{21} + a_{m1}a_{2I})$, where K/L is the endowment ratio of the relatively capital abundant country, even though $a_{11}/a_{21} > K/L > a_{1I}/a_{2I}$ is assumed to hold, as shown in Figure 1.

If the total capital/labor ratio used in the production of final commodity 1 is smaller in the capital abundant country than the endowment ratio, it follows that the intermediate product will be imported rather than exported by the capital abundant country. But, on the basis of the Heckscher-Ohlin theorem, one would expect that the relatively capital abundant country would export the intermediate product, because it is more capital intensive (at any given factor price ratio) than final commodity $2(a_{1I}/a_{2I} > a_{12}/a_{22})$.

It would therefore seem that if the nonspecialization assumption is dropped, the Heckscher-Ohlin theorem is rehabilitated in the sense that it would seem to be suitable not only to predict the trading pattern when trade is confined to final commodities, but also in the more general case, when trade also takes place in intermediate products (provided that the endowment ratios of both countries lie between the total primary factor intensity ratios used in the production of the two final commodities).

Unfortunately, this conjecture is not correct, as may be shown again by means of Figure 1. For this purpose, it is convenient to change the labelling of Figure 1 as follows: the intermediate product is now the most capital intensive commodity and final commodity 1 is less capital intensive than the intermediate product, but more capital intensive than final commodity 2. All this amounts to is to exchange the symbols q_1 and p_I on Figure 1. It is again assumed that over the relevant range of factor price ratios p_2/p_1, no factor intensity reversals occur, i.e., any pair of cost functions has at most one point of intersection. Again, assume that country 1 is in equilibrium at A and country 2 at B. Both countries now produce commodity 1 and country 1 also produces the intermediate product which it definitely will be exporting to country 2. Country 2 produces final commodities 1 and 2 and, given the assumption of nonspecialization in consumption, it follows that country 2 will export commodity 2.

Clearly, the Heckscher-Ohlin theorem is successful in so far as the capital abundant economy exports the most capital intensive product and the more labor abundant economy exports the most labor intensive commodity. The problem is only whether, on the basis of the Heckscher-Ohlin theorem, one can predict whether final commodity 1 is going to be exported or imported by country 1. It appears that the Heckscher-Ohlin theorem is not capable of answering this question. Even if it is assumed as usual that demand patterns are identical between countries and the proportions in which final commodities are demanded are independent of income levels (which are bound to be different in the present example because primary factor prices are not equalized by trade and, of course, factor endowments are different), there is no reason why the labor abundant economy should not export the final commodity, which is relatively more capital intensive (relative to the other final commodity).

To summarize the results of Section I: if it is assumed that no factor intensity reversals occur between any pair of the three products, the Heckscher-Ohlin theorem is completely successful in predicting the trading pattern only if the capital/labor ratio of the intermediate product is nonextremal in the sense that it lies between the direct capital/labor ratios of the final commodities and if the two factor endowments lie between the total primary factor intensity ratios used in the production of both final commodities.

II. Pure Traded Intermediate Products and the Traditional Heckscher-Ohlin Theorem

The main assumptions of the analysis that follows are:

1: The home and foreign countries both produce one final commodity and one pure intermediate product in the original pretrade equilibrium under conditions of universal free trade.

2: The final commodity is considered to be relatively capital intensive and there are two primary factors of production which are always fully employed.

The model is as follows:

a) Factor market equilibrium conditions:

$$a_{11}C_1 + a_{1I}C_I = K$$

$$a_{21}C_1 + a_{2I}C_I = L$$

a_{ij} = primary factor input coefficients (where $j = 1, I$)

$K; L$ = factor endowments,

C_1 = supply of final commodity

C_I = supply of intermediate product

The final commodity is capital intensive and this implies:

$$\begin{vmatrix} a_{11} & a_{1I} \\ a_{21} & a_{2I} \end{vmatrix} > 0$$

b) Cost functions:

$$a_{11}p_1 + a_{21}p_2 + a_{m1}p_I = q_1$$

$$a_{1I}p_1 + a_{2I}p_2 = p_I$$

$p_1; p_2$ = primary factor prices

p_I = price of the intermediate product

q_1 = price of the final commodity
q_1 = numeraire

c) Product market equilibrium condition:

$a_{m1}C_1 = C_I$
a_{m1} = the produced factor input coefficient

The set of equations a), b), and c) represents a closed economy. One wants to find out whether an exogenous increase in $K(L=L_0)$ will lead to a rise or a fall in the price of p_I, depending upon the sign of the determinant

$$\begin{vmatrix} a_{11} & a_{1I} \\ a_{21} & a_{2I} \end{vmatrix}$$

A. *A Very Special Case: Fixed Intermediate Input Coefficients*

If intermediate input coefficients are fixed, the Heckscher-Ohlin theorem can be proved without any difficulty at all. It can easily be seen from an inspection of the set of the factor market equilibrium conditions and cost functions that the following homogeneous, differentiable supply functions, defined for a convex production possibility set $S(C_1, C_I)=0$ with values added and factor endowments as their arguments, are implied:

$$C_1 = C_1(q_1 - a_{m1}p_I, p_I, K, L)$$
$$C_I = C_I(q_1 - a_{m1}p_I, p_I, K, L)$$

Therefore:

$$C_I(q_1 - a_{m1}p_I, p_I, K, L) = a_{m1}C_1(q_1 - a_{m1}p_I, p_I, K, L)$$

The model can be reduced to one equation in one variable p_I. In order to establish the validity of the Heckscher-Ohlin theorem, the last equation is differentiated totally with respect to K.

Solving for dp_I/dK, one obtains:

$$\frac{dp_I}{dK} = \frac{\dfrac{\partial C_1}{\partial K}a_{m1} - \dfrac{\partial C_I}{\partial K}}{\dfrac{\partial C_I}{\partial p_I} - \left(a_{m1}\dfrac{\partial C_I}{\partial pv_1} + a_{m1}\dfrac{\partial C_1}{\partial p_I}\right) + a_{m1}^2\dfrac{\partial C_1}{\partial pv_1}}$$

where $pv_1 = q_1 - a_{m1}p_I$

From the fact that national income may be written as a convex function of values-added $(q_1 - a_{m1}p_I, p_I)$, it follows that the denominator is positive. It can also be readily ascertained that the Rybczynski theorem must hold in the present model, i.e., an increase in the capital endowment (for fixed prices) will bring about an increase in the output of the capital intensive commodity and a decrease in the output of the other product. From

$$\begin{vmatrix} a_{11} & a_{1I} \\ a_{21} & a_{2I} \end{vmatrix} > 0$$

it follows that $\partial C_1/\partial K > 0$ and $\partial C_I/\partial K < 0$, so that the economy with the greater capital endowment will definitely have a comparative advantage in whichever product is relatively capital intensive.

B. *The Assumption of Fixed Intermediate Input Coefficients Relaxed*

The problem now becomes considerably more complicated. The method of solution is to derive three equations in the following three variables: $d\ log\ C_1/d\ log\ K$; $d\ log\ C_2/\ d\ log\ K$; and $d\ log\ p_I/d\ log\ K$ from the set of equations under a), b), and c) given above.

Differentiating first totally the cost functions with respect to K, one obtains:

$$\alpha_{11}\frac{d\log p_1}{d\log K} + \alpha_{21}\frac{d\log p_2}{d\log K} = -\alpha_{m1}\frac{d\log p_I}{d\log K}$$

$$\alpha_{1I}\frac{d\log p_1}{d\log K} + \alpha_{2I}\frac{d\log p_2}{d\log K} = \frac{d\log p_I}{d\log K}$$

$$\text{where } \alpha_{ij} = \frac{a_{ij}p_i}{\text{product price}}$$

Solving, we obtain:

$$\frac{d\,log\,p_1}{d\,log\,K} = \frac{(-\alpha_{m1}\alpha_{2I} - \alpha_{21})}{|\alpha|}\,\frac{d\,log\,p_I}{d\,log\,K}$$

and

$$\frac{d\,log\,p_2}{d\,log\,K} = \frac{(\alpha_{11} + \alpha_{1I}\alpha_{m1})}{|\alpha|}\,\frac{d\,log\,p_I}{d\,log\,K}$$

where $|\alpha|$ is the determinant of cost share ratios. Further below let $a_1 = (-\alpha_{m1}\alpha_{2I} - \alpha_{21})/|\alpha|$ and $a_2 = (\alpha_{11} + \alpha_{1I}\alpha_{m1})/|\alpha|$. Differentiating totally the factor market equilibrium conditions with respect to K, we obtain:

$$b_{11}\frac{d\,log\,C_1}{d\,log\,K} + b_{1I}\frac{d\,log\,C_I}{d\,log\,K} = 1 - b_{11}\frac{d\,log\,a_{11}}{d\,log\,K} - b_{1I}\frac{d\,log\,a_{1I}}{d\,log\,K}$$

$$b_{21}\frac{d\,log\,C_1}{d\,log\,K} + b_{2I}\frac{d\,log\,C_I}{d\,log\,K} = - b_{21}\frac{d\,log\,a_{21}}{d\,log\,K} - b_{2I}\frac{d\,log\,a_{2I}}{d\,log\,K}$$

where b_{ij} = original factor allocation ratios, for example, $b_{11} = a_{11}C_1/K$. Now

$$\frac{d\,log\,a_{11}}{\partial\,log\,K} = \frac{\partial\,log\,a_{11}}{\partial\,log\,p_1}\,\frac{d\,log\,p_1}{d\,log\,K} + \frac{\partial\,log\,a_{11}}{\partial\,log\,p_2}\,\frac{d\,log\,p_2}{d\,log\,K} + \frac{\partial\,log\,a_{11}}{\partial\,log\,p_I}\,\frac{d\,log\,p_I}{d\,log\,K}$$

$$\frac{d\,log\,a_{1I}}{d\,log\,K} = \frac{\partial\,log\,a_{1I}}{\partial\,log\,p_1}\,\frac{d\,log\,p_1}{d\,log\,K} + \frac{\partial\,log\,a_{1I}}{\partial\,log\,p_2}\,\frac{d\,log\,p_2}{d\,log\,K}$$

and now let

$$b_{11}\frac{d\,log\,a_{11}}{d\,log\,K} + b_{1I}\frac{d\,log\,a_{1I}}{d\,log\,K} = A_{11}\frac{d\,log\,p_1}{d\,log\,K} + A_{12}\frac{d\,log\,p_2}{d\,log\,K} + A_{1I}\frac{d\,log\,p_I}{d\,log\,K}$$

where

$$A_{11} = b_{11}\frac{\partial\,log\,a_{11}}{\partial\,log\,p_1} + b_{1I}\frac{\partial\,log\,a_{1I}}{\partial\,log\,p_1} < 0$$

$$A_{12} = b_{11}\frac{\partial\,log\,a_{11}}{\partial\,log\,p_2} + b_{1I}\frac{\partial\,log\,a_{1I}}{\partial\,log\,p_2} > 0$$

(because complementarities between factors are ruled out by assumption) and

$$A_{1I} = b_{11}\frac{\partial\,log\,a_{11}}{\partial\,log\,p_I} > 0$$

Making use of these new definitions and the expressions for $d\,log\,p_1/d\,log\,K$ and $d\,log\,p_2/d\,log\,K$, the differentiated factor market equilibrium conditions may be rewritten as:

$$(3)\quad b_{11}\frac{d\,log\,C_1}{d\,log\,K} + b_{1I}\frac{d\,log\,C_I}{d\,log\,K} + (A_{11}a_1 + A_{12}a_2 + A_{1I})\frac{d\,log\,p_I}{d\,log\,K} = 1$$

$$(4)\quad b_{21}\frac{d\,log\,C_1}{d\,log\,K} + b_{2I}\frac{d\,log\,C_I}{d\,log\,K} + (A_{21}a_1 + A_{22}a_2 + A_{21})\frac{d\,log\,p_I}{d\,log\,K} = 0$$

where A_{21}, A_{22}, and A_{2I} are defined in complete analogy to A_{11}, A_{12}, and A_{1I}, and we know that $A_{21} > 0$, $A_{22} < 0$, and $A_{2I} > 0$.

The third equation is obtained by totally differentiating the product market equilibrium condition:

$$\frac{d\,log\,C_1}{d\,log\,K} - \frac{d\,log\,C_I}{d\,log\,K} + \left(\frac{\partial\,log\,a_{m1}}{\partial\,log\,p_1}a_1 + \frac{\partial\,log\,a_{m1}}{\partial\,log\,p_2}a_2 + \frac{\partial\,log\,a_{m1}}{\partial\,log\,p_I}\right)\frac{d\,log\,p_I}{d\,log\,K} = 0$$

Solving for $d\,log\,p_I/d\,log\,K$, one obtains for the numerator: -1 and for the denominator:

$$a_1\Big(A_{21} - A_{11}\Big) + a_2\Big(A_{22} - A_{12}\Big) +$$

$$A_{2I} - A_{1I} + \left(\frac{\partial \log a_{m1}}{\partial \log p_1} a_1 + \frac{\partial \log a_{m1}}{\partial \log p_2} a_2 + \frac{\partial \log a_{m1}}{\partial \log p_I}\right) |b|$$

The problem now is to find out whether a definite sign can be derived for $d \log p_I / d \log K$, given that $|\alpha| > 0$ (the final commodity is capital intensive) and given that $a_1 < 0$, $a_2 > 0$, and $A_{11} < 0$, $A_{22} < 0$, but $A_{12} > 0$, $A_{21} > 0$, and $A_{1I} > 0$ and $A_{2I} > 0$. Also, $\partial \log a_{m1} / \partial \log p_1 > 0$ and $\partial \log a_{m1} / \partial \log p_2 > 0$, but $\partial \log a_{m1} / \partial \log p_I < 0$. It should also be noted that $A_{11} + A_{12} + A_{1I} = 0$ and $A_{21} + A_{22} + A_{2I} = 0$ (from the homogeneity property of input coefficients). Similarly, it is known that

$$\frac{\partial \log a_{m1}}{\partial \log p_I} = -\frac{\partial \log a_{m1}}{\partial \log p_1} - \frac{\partial \log a_{m1}}{\partial \log p_2}$$

Making use of these expressions the denominator of $d \log p_I / d \log K$ becomes:

$$\left(A_{21} - A_{11}\right)(a_1 - 1) + \left(A_{22} - A_{12}\right)(a_2 - 1) + \left[\frac{\partial \log a_{m1}}{\partial \log p_1}(a_1 - 1) + \frac{\partial \log a_{m1}}{\partial \log p_2}(a_2 - 1)\right] |b|$$

Furthermore, it can easily be shown that $|\alpha| > 0$ implies that $a_1 < 0$, $a_2 > 1$, and also $|b| > 0$.

For the Heckscher-Ohlin theorem to remain unscathed, clearly the denominator would have to be negative. If there are no complementarities between factors, the expression $(A_{21} - A_{11})(a_1 - 1) + (A_{22} - A_{12}) \cdot (a_2 - 1)$ will be negative.[2] But unfortunately the sign of

$$(5) \quad \left[\frac{\partial \log a_{m1}}{\partial \log p_1}(a_1 - 1) + \frac{\partial \log a_{m1}}{\partial \log p_2}(a_2 - 1)\right]$$

is unknown. It therefore follows that the Heckscher-Ohlin theorem breaks down. As a considerable amount of trade takes place in intermediate products, this seems to be a result of some importance. The Heckscher-Ohlin theorem is inadequate in explaining trading patterns when pure traded intermediate products are allowed for. There are the following special cases in which the Heckscher-Ohlin theorem remains valid:

1) when intermediate input coefficients are fixed

2) when $\dfrac{d \log a_{m1}}{d \log p_I} = \dfrac{\partial \log a_{m1}}{\partial \log p_I}$

It may easily be shown that if condition 2) is fulfilled, expression (5) vanishes. The economic interpretation of the second set of special conditions is quite straightforward. The Heckscher-Ohlin theorem remains valid if the total elasticity of the intermediate input coefficient (allowing for variations in the prices of original factors of production) is equal to the partial elasticity of the intermediate input coefficient (keeping p_1 and p_2 fixed) with respect to the price of the intermediate product. Briefly, the condition under 2) can be interpreted to entail that substitution between original factors of production and the intermediate input is unbiased.[3]

REFERENCES

R. N. Batra and F. R. Casas, "Intermediate Products and the Pure Theory of International Trade: A Neo-Heckscher-Ohlin Framework," *Amer. Econ. Rev.*, June 1973, *63*, 297–311.

---

[2] If complementarities between factors are allowed for, the sign of $(A_{21} - A_{11})(a_1 - 1) + (A_{22} - A_{12})(a_2 - 1)$ becomes unknown.

[3] On this concept see the book by W. M. Corden.

W. M. Corden, *The Theory of Protection,* Oxford 1971.

R. W. Jones, "Effective Protection and Substitution," *J. Int. Econ.,* Feb. 1971, *1,* 59–81.

J. S. Kelly, "Lancaster vs. Samuelson on the Shape of the Neo-classical Transformation Surface," *J. Econ. Theory,* Oct. 1969, *1,* 347–51.

M. C. Kemp, *The Pure Theory of International Trade and Investment,* Englewood Cliffs 1969.

J. R. Melvin, "Production and Trade with Two Factors and Three Goods," *Amer. Econ. Rev.,* Dec. 1968, *58,* 1249–68.

A. G. Schweinberger, "The Heckscher-Ohlin Model and Traded Intermediate Products," *Rev. Econ. Stud.,* Apr. 1975, *42,* 269–78.

S. Vanek, "Variable Factor Proportions and Inter-Industry Flows in the Theory of International Trade," *Quart. J. Econ.,* Feb. 1963, *77,* 129–42.

R. D. Warne, "Intermediate Goods in International Trade with Variable Proportions and Two Primary Inputs," *Quart. J. Econ.,* May 1971, *85,* 225–36.

Did the 1968 Surcharge Really Work?

By William L. Springer*

In a recent article, Arthur Okun (1971) has examined the effects on aggregate consumer expenditures of the recent temporary income tax surcharge (officially known as the Revenue and Expenditure Act of 1968), using the consumption sector equations from four prominent macroeconomic models (Data Resources, Inc. (*DRI*), Wharton, Office of Business Economics, and Michigan). On the basis of his analysis, Okun concluded that "the evidence of the surcharge period as interpreted by four econometric models indicates that the surcharge curbed consumption nearly as much as was expected, and that any shortcomings in its effectiveness have no evident connection to the permanent income hypothesis" (p. 198). Moreover, "the evidence . . . provides further confirmation of the general efficacy and continued desirability of flexible changes in personal income tax rates—upward or downward, permanent or temporary" (p. 200). Thus, in spite of its admitted failure to halt the inflationary surge of the late 1960's, Okun claims that the surcharge accomplished what it was designed to do—namely, to reduce consumer expenditures. The persistence of inflation would have to be explained by (perhaps) offsetting increases in other components of aggregate demand (for example, state and local government expenditures), an overexpansive monetary and credit policy, inflationary expectations, or cost-push pressures from big business and labor.[1]

Viewed against the widely accepted "permanent income" and "life cycle" hypotheses of consumer behavior which predict small changes in consumption in response to temporary changes in income, Okun's results are somewhat puzzling. As Robert Eisner (1969) has pointed out, " . . . the tax surcharge should never, on basic theoretical grounds, have been considered an effective anti-inflationary device. . . . The basic economic error of those who saw in the 10 percent income tax surcharge an adequate measure against inflation may be charged to failure to take into account the implications of the permanent income hypothesis . . . " (p. 898). The question, then, is how can Okun's apparent finding that the surtax "worked" be reconciled with the prior expectation that it would not work because of its temporary nature.

As part of a more extended study of windfall income, temporary income taxes, and consumer behavior, I have performed essentially the same kind of error analysis over the surcharge period as Okun did. However, using more suitable estimation and simulation procedures, along with a somewhat different set of consumption equations, I arrived at exactly the opposite conclusions. In particular, it was found

---

* Senior economist, Data Resources, Inc. This paper is based on two chapters of my unpublished doctoral dissertation. I wish to thank Alan Blinder and William Branson of Princeton University and an anonymous referee for extremely helpful comments and suggestions, but I alone assume responsibility for any remaining errors.

[1] Using simple Keynesian macro models, Alan Blinder has investigated the conditions under which an increase in income taxes might actually lead to *higher* prices (via the indirect route of workers demanding a constant after-tax wage). Except in a Monetarist model, the effect of such a tax change on prices is always ambiguous. However, Blinder concludes that under reasonable assumptions concerning values for the marginal tax rate, the tax multiplier, labor's share, and the supply and demand elasticities for labor, income tax increases will have the desired deflationary impact.

that the 1968 surcharge did *not* lead to a significant reduction in consumption expenditures. This result is consistent not only with generally accepted theory, but also with the well-known fact that a significant drop in the savings rate occurred in late 1968 and persisted throughout 1969. The finding that consumer expenditures were not effectively curtailed by the tax also helps explain why inflation continued unabated after 1968.

It therefore appears that this episode of fiscal policy does not in fact provide any evidence whatsoever that temporary income taxes will have a discernible impact on consumption. On the contrary, the data seem to confirm what theory predicts and what casual empiricism at the time suggested. These findings should be of particular interest to government fiscal policy makers who, within a recent period of about four months, proposed first a temporary 5 percent income tax surcharge and then a one-time rebate on 1974 taxes. Such small and temporary measures simply cannot be expected to have a significant effect on total aggregate demand.

Section I lays some rudimentary theoretical groundwork, Section II discusses the nature of the experiment, Section III describes the data, Section IV presents the empirical results, and Section V summarizes the main conclusions.

I. Theoretical Considerations

Although Okun's approach suffers from its lack of grounding in a well-developed theoretical framework, the intention here is not to present a full-blown treatment of windfalls, but rather to challenge Okun on his own ground, i.e., by performing experiments similar to those reported in his paper.[2] Nevertheless, since my equations do follow from specific theoretical models, it is necessary to outline briefly their derivation.

The first point to be noted is that my equations split consumer expenditures according to the economist's rather than the national income accountant's definitions. That is, "consumption" is here defined to include depreciation and imputed interest on durables, as well as expenditures on nondurables and services. This might seem to muddy the waters a bit, since Okun's equations are all based on National Income Accounts (*NIA*) concepts. Moreover, from a policymaker's point of view, the ultimate concern is with *expenditures* rather than the theoretically more appealing concept of consumption. However, the difference turns out to be more apparent than real, since expenditures on nondurables and services constitute roughly 80 percent of total consumption, and the "use value" on durables is essentially a trend. Thus, movements in consumption will be dominated by movements in expenditures on nondurables and services. As for consumer durables, the model can be developed either in terms of net or gross expenditures. The latter concept was chosen for this study and is also the one used in Okun's forecasting equations.

A. *Consumption*

The consumption equation follows from the straightforward maximization of a well-behaved, additively separable, intertemporal utility function, subject to the budget constraint that discounted consumption plus bequests must equal the sum of discounted future disposable income.[3] Assuming the utility function to be homogeneous of some positive degree, the maximization process yields an optimal consumption path which is a proportional function of lifetime income, where the proportionality factor depends on an individual's age, his subjective rate of time preference (tastes), and the market rate of interest.

---

[2] A theory of windfalls is developed in my dissertation, together with a discussion of its application to the question of flexible tax policy.

[3] In most studies of the consumption function, the budget constraint is expressed in terms of initial assets and *labor* income. Because of the difficulty in obtaining an accurate and consistent series for labor income, lifetime resources are here defined as the sum of the discounted *disposable* income stream. Since this concept includes income from property, net worth should not appear as an explanatory variable in the consumption function.

$$(1)\qquad con^*_{it} = k_i\left[\sum_{t=T}^{L} \frac{yd^e_{it}}{(1+rr_t)^{t-T+1}}\right]$$

where

con^*_{it} = desired consumption of individual i in time period t
yd^e_{it} = disposable income expected (with certainty) in time period t
rr_t = real market rate of interest
T = present age of individual i
L = life span of individual i
k_i = proportionality factor
= k (age, tastes, rate of interest)

In the spirit of Franco Modigliani and his collaborators (1954, 1963) and Milton Friedman, lifetime disposable income is converted into a yearly income flow—"annual expected disposable income" ($\overline{yd}^e_{it}$)—defined as the implicit (constant) annuitized income stream. Thus,

$$(2)\qquad \overline{yd}^e_{it} = \frac{\displaystyle\sum_{t=T}^{L} \frac{yd^e_{it}}{(1+rr_t)^{t-T+1}}}{\displaystyle\sum_{t=T}^{L} \frac{1}{(1+rr_t)^{t-T+1}}}$$

$$(3)\qquad con^*_{it} = \alpha_i(\overline{yd}^e_{it})$$

where[4]

$$\alpha_i \equiv k_i \sum_{t=T}^{L} \frac{1}{(1+rr_t)^{t-T+1}}$$

Under the assumption of a constant proportionality factor α for each age group, a stable population distribution, and a stationary income distribution when (3) is aggregated over the entire population, the corresponding macro coefficient will be time invariant. Whether it differs at a given point in time from the simple average of α_i depends on the covariance between these micro slopes and the income distribution parameters. If, for example, the marginal propensity to consume (*MPC*) declines with income as many have claimed, the estimated macro coefficient will exhibit negative "aggregation bias," i.e., the macro *MPC* will understate the unweighted mean of the micro *MPC*s. But this aggregation bias (if it exists) at least will not vary from period to period under the stated assumptions.

Aggregated desired consumption in each period is now:

$$(4)\qquad CON^*_t = \beta\overline{YD}^e_t$$

where

CON^*_t = aggregate desired consumption
$\overline{YD}^e_t$ = aggregate annual expected disposable income
$\beta = \beta$ (age, tastes, rate of interest)

Aggregate *measured* consumption is the sum of desired and transitory consumption.

$$(5)\qquad CON_t = CON^*_t + CON^r_t$$

where

CON_t = measured consumption
CON^*_t = desired or permanent consumption
CON^r_t = residual or transitory consumption

If we depart from the strict Friedman hypothesis of zero correlation between transitory income and consumption, and allow for some expansion and contraction of consumption (for example, of nonessential items) in response to transitory income changes,[5] then aggregate transitory consumption is defined as:

---

[4] Under the assumptions of an infinite lifetime and a constant rate of interest, $\overline{yd}^e_{it}$ becomes identical to Friedman's "permanent income."

[5] Because the "consumption" variable used here is constructed using *NIA* concepts, it is impossible to ex-

$$CON^T_t = \gamma YT_t \tag{6}$$

where YT_t = aggregate transitory income
$\gamma = MPC$ out of transitory income

This yields for measured consumption:

$$CON_t = \beta \overline{YD}^e_t + \gamma YT_t \tag{7}$$

Using the identity that measured disposable income is the sum of annual expected income and transitory income ($YD_t = \overline{YD}^e_t + YT_t$), the expression for aggregate measured consumption becomes:

$$CON_t = \gamma YD_t + (\beta - \gamma)\overline{YD}^e_t \tag{8}$$

It will be recalled that β is a function of age, tastes, and the rate of interest. While it is probably not too far from the mark to assume that tastes and the age distribution have been relatively stable over the last twenty years, it is less defensible to make such an assumption with respect to the interest rate. In the early studies of the consumption function, the interest rate was assumed to be zero or constant in time. At the time these articles appeared, rates of interest showed relatively little fluctuation, so that such assumptions may have been legitimate. However, in recent years, both real and nominal rates have varied significantly, so that the potential impact on consumption can no longer be safely ignored.

In principle, β is a function of current and future rates of interest. However, the simplifying assumption will be made that it is a linear function of the current rate only, i.e., that

$$\beta = b_0 + b_1 rr_t \tag{9}$$

Substituting this into (8), we have:

$$CON_t = \gamma YD_t + (b_0 + b_1 rr_t - \gamma)\overline{YD}^e_t \tag{10}$$

Or, simplifying the notation, the final estimating equation for consumption is:

$$CON_t = B_{11} YD_t + B_{12}\overline{YD}^e_t + B_{13}(rr_t \cdot \overline{YD}^e_t) + u_t \tag{11}$$

where u_t is the stochastic error term.

B. *Consumer Durables*

The model of investment in consumer durables used in this study draws heavily on the work of Thomas Juster and Paul Wachtel (1972a,b) and Saul Hymans. Basically, their aim is to supplement the traditional "objective" stock-adjustment model of durables expenditures with "anticipatory" variables designed to measure consumer sentiment. They find that these sentiment indexes do indeed improve the explanatory power of durables equations. Without going through the theoretical derivation, the estimating equation for gross expenditures on consumer durables looks as follows:

$$\begin{aligned} CD_t = {} & B_{21} + B_{22} YD_t + B_{23}\overline{YD}^e_t \\ & + B_{24}\left(\frac{PCD}{PNS}\right)_t + B_{25} rr_t \\ & + B_{26} KCD_{t-1} + B_{27} UN_t \\ & + B_{28} ASTRIKE + B_{29} NCD_{t-1} \\ & + B_{210} UN_{t-1} + B_{211} FMOOD_t + v_t \end{aligned} \tag{12}$$

where CD_t = gross expenditures on consumer durables
NCD_t = net expenditures on consumer durables
KCD_t = net stock of consumer durables
PCD_t = implicit deflator for durables
PNS_t = implicit deflator for nondurables and services
YD_t = current disposable income
$\overline{YD}^e_t$ = average annual expected income
rr_t = real rate of interest

---

clude discretionary expenditure terms entirely. Thus, while in principle it is possible to imagine some subset of total consumption expenditures which is totally uncorrelated with transitory income, in practice this is difficult to attain because of the way the data are actually measured and reported.

UN_t = aggregate unemployment rate

$ASTRIKE_t$ = automobile strike dummy: +1 in 59-IV, 64-IV, 67-IV and 70-IV; and −1 in 60-I, 65-I, 68-I and 71-I

$FMOOD_t$ = "filtered" version of the Michigan Survey Research Center's Index of Consumer Sentiment.[6]

II. The Nature of the Experiment

Okun's experiment consisted of measuring the prediction errors of the consumption sector equations under two different assumptions. According to the "full-effect" hypothesis, the surtax was treated like any other change in disposable income, so that consumption over the period would be best explained by actual measured disposable income. The alternative "zero-effect" view holds that consumption was closer to what would have been predicted had there been no surcharge at all. That is, this hypothesis essentially says that consumers behaved as if they had lost no income at all. Thus, in this case, the relevant explanatory variable is disposable income, less the (negative) surtax, i.e., the amount of the tax is added back. Our prior expectation is that actual consumption should fall somewhere in the interval bounded by predictions based on the two extreme views. If the residuals have a positive bias on the full-effect view and no bias on the zero-effect view, the latter hypothesis would be confirmed. But if they have a negative bias on the zero-effect view and are unbiased on the full-effect view, the full-effect hypothesis would be validated.[7]

The prediction errors of my equations will be compared over the entire period 1968-III to 1971-IV and for the subperiods, 1968-III to 1969-II, 1968-III to 1970-II, and 1970-III to 1971-IV. There are two reasons for making comparisons over these particular intervals. In the first place, from the point of view of the effectiveness of stabilization policy, the major concern is with what happens in the first and second years following imposition of the surcharge. For if temporary taxes and transfers are seriously being considered as a means of stabilizing aggregate demand, then it is certainly to be hoped that they will yield significant results within at least two years, and preferably within one. It therefore seemed appropriate to examine the residuals over shorter periods of time following initiation of the tax. (Okun looked only at the nine-quarter interval, 1968-III to 1970-III.)

A second and more practical reason for

---

[6] See Juster and Wachtel (1972b) for a description of this variable.

[7] While Okun discusses the permanent income arguments against effectiveness of the surtax and even claims to have refuted them with his results, his experiment does not in fact provide a direct test of the theory. In the first place, by using an equation estimated with actual disposable income and by adjusting only the current income term, the experiment *assumes* that the surcharge was treated like any other change in disposable income. But this is the central point at issue. Secondly, validation of the full-effect view does not necessarily imply contradiction of the permanent income hypothesis, for the simple reason that nearly all the equations used by Okun implicitly assume this theory by incorporating lagged values of consumption and income. Moreover, most of his equations carry very small coefficients on the current income term—a result which is predicted by the permanent income hypothesis and which suggests immediately that temporary changes in income will have a small impact on consumption. And finally, the zero-effect view in effect assumes that the estimated historical relationship between consumer expenditures and income was somehow altered during the surcharge period, so that consumers behaved *as if* they had more income than they really did. Thus, the amount of the tax is arbitrarily added back to disposable income to force the equation to predict a higher level of consumption. But the permanent income theorists do not need to assume that the surtax did not happen in order to obtain their prediction of ineffectiveness. Their point is simply that because the income loss was temporary and because such changes in income carry a small *MPC*, consumption was little affected. A direct test of the permanent income hypothesis requires an examination of the distributed lag behavior of a separately entered surcharge variable and estimation of the propensity to consume out of it. Such a test is performed in the author's dissertation.

splitting out the last six quarters and treating them separately is that there seems to have been some factor, not fully accounted for in the specified equation, which caused consumption to be unusually *low* in 1971. It seems doubtful that this was related to the surcharge, and even if it was, it does not say much for the short-run effectiveness of the measure. Moreover, this analysis is really concerned with first-round effects of the tax, and these were over by 1970-IV. Thus, since 1971 seems to confuse the issue somewhat, results which exclude this period will be presented.

Three measures of error are used in analyzing the residuals: mean absolute error (MAE), root mean squared error ($RMSE$), and the mean error (ME). Note that the residual $\hat{e}$ used in the calculation of these measures is the nonautocorrelated error term.[8]

In calculating predicted values of the expenditure variables, lagged residuals are generated by using lagged *predicted* values of the dependent variables. This "dynamic" procedure is generally considered to be the acid test of the predictive accuracy of an equation containing the lagged value of the dependent variable and/or an autocorrelated error term. If instead, lagged residuals are generated using *actual* lagged values of the dependent variable, then as stated by Okun: "predictions based on an incorrect theoretical view of consumption are mechanically corrected and approach the predictions based on the valid analytical view" (1971, p. 182). On the other hand, under the dynamic method, if predictions are consistently wrong in one direction or another, they will cumulate and become very obvious. An equation which predicts well when errors are fed back into it dynamically can with more accuracy be said to reflect the valid analytical view.

III. Data and Estimation Methods

The data used in estimating equations (11) and (12) are expressed in billions of 1958 dollars seasonally adjusted at annual rates, and are derived from the standard National Income Accounts tables published by the Department of Commerce. The sample period runs from the first quarter of 1955 to the fourth quarter of 1971.

Consumption (CON) is defined as the sum of expenditures on nondurable goods and services (CNS), depreciation on the stock of durables (WCD), and the implicit stream of rental services on the existing stock of durables (YCD).[9] Since consumption includes the implicit yield on durables, YCD must be added to measured disposable income in order to maintain the accounting identity. Data on the estimated dollar amounts of tax accruals yielded by the surcharge were supplied by the Department of Commerce and may be found in Okun (1971).[10] In all cases, the expected income and interest rate interaction terms have been made consistent with the definition of current disposable income.

---

[8] Ideally, confidence intervals should also be constructed around the prediction errors and significance tests performed thereon. This is not done because the additional computation involved was not readily available. Furthermore, it is highly unlikely that such tests would alter the nature of the results derived from the simple error measures used here.

[9] See William Poole for a convenient exposition of the method by which CON is constructed.

[10] Albert Ando and E. Cary Brown discuss the question as to whether consumer expenditures in the short run can best be explained by disposable income with taxes subtracted on a liability basis, or an accrual basis (the latter concept being the one used in the National Income Accounts). Under normal assumptions, a utility optimum is achieved with a smooth consumption path, so that we would expect rational individuals to adjust their consumption patterns to the smooth trend of tax liabilities rather than to tax accruals, which *may* be more volatile. Actually, except in unusual circumstances, accruals and liabilities are likely to move closely together and therefore should perform equally well in explaining consumer expenditures. It is only when they exhibit sharply differing movements that we must be concerned about which concept to use in a consumption equation. There is no evidence that such a divergence occurred over the surcharge period.

Annual expected income is not observable and so must be approximated in some manner. The most common way of doing this is to assume that expected disposable income is generated according to the adaptive expectations mechanism suggested by Marc Nerlove. Thus,

$$(13)\quad \begin{aligned}\overline{YD}^e_t &= \overline{YD}^e_{t-1}(1+g) \\ &\quad + h(YD_{t-1} - \overline{YD}^e_{t-1}) \\ &= hYD_{t-1} + (1 - h + g)\overline{YD}^e_{t-1}\end{aligned}$$

where g is a quarterly growth trend factor and h is the coefficient of expectation or adjustment, which takes on a value between zero and one. The initial value for $\overline{YD}^e_{t-1}$ (fourth quarter of 1948) was calculated as the simple average of disposable income from 1948-II through 1949-II, or \$232.8 billion. The growth rate of income (.0094) was the estimated slope in a regression of the *log* of disposable income on a constant and time. A value of .1 for h was chosen as that value between .1 and .2 which minimized the sum of squared residuals in a series of regressions using equation (11). The same procedure was followed in calculating $\overline{YD1}^e$, average expected disposable income, less the (negative) surcharge.

The nominal interest rate rn is a weighted average of rates on savings and loan shares, mutual savings bank deposits, and time deposits in commercial banks with weights being determined by the share of each asset in the portfolio of savings deposits in the household Flow of Funds Accounts as of the end of 1971. The nominal rate of interest rn was converted to a real rate of interest rr by subtracting the expected rate of inflation PE. Both rn and rr are expressed in percent.

PE was calculated by assuming an adaptive expectations mechanism for price formation.

$$(14)\quad PE_t = PE_{t-1} + \bar{h}(dP_t - PE_{t-1})$$

where PE_t = expected rate of price change (percent)

dP_t = actual percentage rate of price change $= ((P_t - P_{t-1})/P_{t-1}) \times 100$

$\bar{h}$ = coefficient of expectation, $0 < \bar{h} < 1$

By recursive substitution, (14) may be written as:

$$(15)\quad PE_t = \bar{h}\sum_{i=0}^{\infty}(1-\bar{h})^i dP_{t-i}$$

Thus, the current expected rate of inflation may be expressed as an infinite geometrically distributed lag on actual past rates of inflation. Rather than estimate the lag structure directly, a set of assumed weights was taken from Robert J. Gordon who estimates an interest rate equation with an Almon polynomial distributed lag on past price changes. It was arbitrarily decided to use the weights for the first three years only (12 quarters), and to normalize them to sum to one.

The price variable used in this work is the implicit GNP deflator. This was chosen to avoid both the possible spurious bias caused by using the consumption deflator in consumption equations, and the apparent upward bias of the fixed-weight Consumer Price Index as noted by Gordon, p. 111.

Although contemporaneous correlation undoubtedly exists between the regressors and the disturbance, so that least squares estimates will be biased and inconsistent, the results reported here have not accounted for the simultaneous equation problem. All equations were estimated by ordinary least squares, using the Cochrane-Orcutt iterative technique to correct for autocorrelated residuals. Ray Fair (1971) has reported that in estimating equations with aggregate quarterly data, it is apparently more important to remove autocorrelation than simultaneity. Two-stage least squares estimates tend to differ little from their ordinary least squares counterparts, whereas the Cochrane-Orcutt transformation often yields quite different coefficient estimates. Some experimentation by this author using a consistent estimation procedure suggested by Fair (1970) corroborated his findings, so that failure to use two-stage least squares estimates has essentially no bearing on the results.[11]

IV. Empirical Results

A. *Consumption*

The equations estimated under the full- and zero-effect views of the surcharge are given below in (16) and (17).

Full-Effect

$$(16)\quad CON_t = \underset{(0.5)}{2.74} + \underset{(4.2)}{0.26}YD_t + \underset{(10.1)}{0.65}\overline{YD}_t^e + \underset{(1.9)}{0.004}(rr_t \cdot \overline{YD}_t^e)$$

$$R^2 = .9996;\quad SE = 1.4664;\quad DW = 2.5270;\quad \hat{\rho} = \underset{(12.6)}{0.84}$$

Zero-Effect

$$(17)\quad CON_t = \underset{(1.0)}{2.84} + \underset{(6.7)}{0.34}YD1_t + \underset{(10.7)}{0.56}\overline{YD1}_t^e + \underset{(4.6)}{0.005}(rr_t \cdot \overline{YD1}_t^e)$$

$$R^2 = .9996;\quad SE = 1.3936;\quad DW = 2.4336;\quad \hat{\rho} = \underset{(7.5)}{0.67}$$

In these equations, YD refers to actual disposable income, while $YD1$ refers to disposable income less the (negative) surcharge. These two variables are identical except for the nine quarters, 1968-III through 1970-III. The t-statistics are shown in parentheses below the coefficient estimates.

A cursory look at the income coefficients in both (16) and (17) should suggest immediately that income taxes of any kind will not be a very effective weapon of stabilization policy. The relatively small estimated coefficients on current income, together with the small coefficient of expectation (.1) on expected income, indicate that even a permanent alteration in income taxes would yield only small short-run changes in consumption, and that the lag between implementation and full impact would be long. And if the tax change were temporary (even assuming it is treated like an ordinary change in disposable income), the long-run impact on consumption would ultimately be much smaller as well. The coefficient on the interest rate interaction term is positive, indicating that the income effect dominates the substitution and asset effects of an interest rate change.

Although the differences in the statistical properties of the two equations are not large, the zero-effect equation (17) is clearly superior to the full-effect equation (16). In (17) the standard error and $\hat{\rho}$ are both smaller, the Durbin-Watson statistic is closer to 2, and the parameters are uniformly more significant. More persuasive evidence of the superiority of the zero-effect view is provided by an examination of the predictive power of the two equations. Tables 1 and 2 present the actual and dynamic residuals for both equations, and the summary measures of their prediction errors over various time periods following institution of the surcharge.

The error pattern indicates that while the percentage errors are quite small, both equations are generally underpredicting consumption in the first eight quarters. However, this is much more pronounced in the case of the full-effect view, where the *RMSE* over this interval for equation (16) is roughly twice that for equation (17).[12]

---

[11] See the author's dissertation for a more extensive discussion of the data and estimation methods used.

[12] Note that the direction of error is just the opposite

TABLE 1—RESIDUALS OF EQUATIONS (16) AND (17)
$(\hat{e}=CON_t-CON_t)$

Quarter	STATIC (ACTUAL) *Full-Effect* (16)	STATIC (ACTUAL) *Zero-Effect* (17)	DYNAMIC *Full-Effect* (16)	DYNAMIC *Zero-Effect* (17)
68-III	3.36	1.95	3.36	1.95
68-IV	0.44	0.24	3.26	1.55
69-I	1.40	0.56	4.13	1.61
69-II	1.99	1.63	5.46	2.71
69-III	−1.05	−0.30	3.53	1.53
69-IV	0.45	0.57	3.41	1.59
70-I	1.74	2.59	4.60	3.66
70-II	−1.94	−2.00	1.93	0.47
70-III	−0.35	0.55	1.27	0.87
70-IV	0.39	1.43	1.46	2.01
71-I	−2.53	−2.36	−1.30	−1.01
71-II	−0.86	−1.05	−1.95	−1.73
71-III	−3.29	−3.43	−4.93	−4.59
71-IV	−0.39	−1.47	−4.53	−4.57

TABLE 2—SUMMARY ERROR MEASURES
(*CON* = Dependent Variable)

Period	DYNAMIC *MAE*[a]	DYNAMIC *RMSE*[b]	DYNAMIC *ME*[c]
A. *Full-Effect View* (16)			
68-III to 69-II	4.052	4.148	4.052
68-III to 70-II	3.710	3.857	3.710
70-III to 71-IV	2.574	3.028	−1.663
68-III to 71-IV	3.223	3.505	1.407
B. *Zero-Effect View* (17)			
68-III to 69-II	1.956	2.062	1.956
68-III to 70-II	1.884	2.062	1.884
70-III to 71-IV	2.464	2.944	−1.505
68-III to 71-IV	2.132	2.478	0.4314

[a] $MAE = \frac{1}{n}\sum_{t=1}^{n} |\hat{e}_t|$

[b] $RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} (\hat{e}_t)^2}$

[c] $ME = \frac{1}{n}\sum_{t=1}^{n} \hat{e}_t$

Over the last six quarters (1970-III to 1971-IV), both equations are overpredicting, but yield residuals of roughly the same size. For the period as a whole (1968-III to 1971-IV), the zero-effect equation yields significantly smaller errors than the full-effect equation, with the difference being most pronounced over the first eight quarters.

Although there is relatively little bias in the residuals over the entire fourteen quarters on the zero-effect view, it is somewhat puzzling that consumption was slightly higher even than predicted by the zero-effect view during the first eight quarters—the period when the surcharge was supposed to be having its major impact—and lower even than predicted by the full-effect view in the last four quarters. It is highly unlikely that the tax *stimulated* consumption while it was in effect and then finally began to *reduce* spending after it was removed. And even if such an hypothesis were not implausible on basic theoretical grounds, it would imply extremely long lags in effectiveness which would render the tax useless as a short-run stabilization device. The fact is that even under the preferred zero-effect view the equation as specified does not track consumption very well in 1971, when something else seems to have been holding down spending. An attempt was made to identify the cause of the unusually low levels of consumption in 1971, but a completely satisfactory answer was not found.[13]

Thus, having performed Okun's experiment using a simple consumption function derived precisely from standard economic theory, it must be concluded that actual consumption was closer to the levels predicted by the zero-effect view than to those predicted by the full-effect view. Contrary to Okun's findings, consumption appears to have been unaffected by the surcharge, as consumers did in fact discount the temporary tax.

---

from that reported by Okun. Except for the *DRI* equations, his models consistently *over*predict consumption of nondurables and services on either view by a disturbingly large margin.

[13] See the author's dissertation for a discussion of these tests.

TABLE 3—RESIDUALS OF EQUATIONS (18) AND (19)
($\hat{e}=CD_t-CD_t$)

	STATIC (ACTUAL)		DYNAMIC	
Quarter	*Full-Effect* (18)	*Zero-Effect* (19)	*Full-Effect* (18)	*Zero-Effect* (19)
68-III	1.75	1.71	1.75	1.71
68-IV	−0.20	0.31	−0.75	−0.73
69-I	1.78	1.60	2.02	1.78
69-II	0.75	0.36	0.12	−0.07
69-III	−1.44	−1.55	−1.48	−1.53
69-IV	−0.02	0.07	0.44	0.45
70-I	0.19	0.33	0.05	0.22
70-II	0.83	0.43	0.81	0.37
70-III	−0.40	−0.56	−0.65	−0.65
70-IV	−3.59	−3.44	−3.38	−3.29
71-I	1.27	1.04	2.34	1.84
71-II	0.13	−0.01	−0.61	−0.46
71-III	1.84	2.20	2.03	2.31
71-IV	0.52	0.84	−0.12	−0.28

TABLE 4—SUMMARY ERROR MEASURES
(CD = Dependent Variable)

		DYNAMIC	
Period	*MAE*[a]	*RMSE*[b]	*ME*[c]
A. *Full-Effect View* (18)			
68-III to 69-II	1.159	1.389	0.785
68-III to 70-II	0.927	1.162	0.371
70-III to 71-IV	1.522	1.909	−0.064
68-III to 71-IV	1.182	1.525	0.184
B. *Zero-Effect View* (19)			
68-III to 69-II	1.072	1.287	0.673
68-III to 70-II	0.858	1.083	0.275
70-III to 71-IV	1.471	1.837	0.006
68-III to 71-IV	1.121	1.453	0.160

[a,b,c] See Table 2.

B. *Consumer Durables*

While it appears that *consumption* was little affected by the income tax surcharge, there is no presumption one way or the other as to the effect on total consumer *expenditures*. The permanent income/life-cycle theory says only that temporary changes in income will be reflected in savings flows, defined to include personal saving (National Income Accounts basis) and net expenditures on durables. Clearly, which component of saving is primarily affected by a temporary tax determines what the ultimate impact on total expenditures will be. Thus, if the saving takes the form of changes in the stock of durable goods, instead of financial assets, then although *consumption* may be unaffected by the tax, the measure must still be considered a successful instrument of stabilization policy since it has caused the desired change in total consumer expenditures.

This is a point often missed by those who have criticized the theoretical justification for the 1968 income tax surcharge. These critics make the valid point that a temporary change in income taxes will probably have a miniscule effect on current consumption, and then conclude that such taxes are a useless tool of stabilization policy. However, the second conclusion does not necessarily follow from the first because they relate to two different concepts—consumption and consumer expenditures. It is the latter concept which is most relevant in the context of economic stabilization, and it is therefore necessary to look into the question of the allocation of temporary income between real and financial assets. On a priori grounds, there is little justification for believing that this allocation occurs in any particular fashion; it would seem to be a purely empirical matter. It is a great irony that it could turn out that the permanent income theorists, who have denigrated the usefulness of temporary income taxes, and the proponents of an active and flexible fiscal policy could both be right!

The durables equation (12) was estimated over the sample period (55-I to 71-IV) using the two definitions of disposable income, and dynamic simulations were run over the surcharge period. In these simulations, only the lagged error term is fed through dynamically. The equations, their residuals, and the summary error measures are shown below.

Full-Effect

$$(18)\quad CD_t = \underset{(0.4)}{8.69} + \underset{(2.3)}{0.08}YD_t + \underset{(0.6)}{0.05}\overline{YD}^e_t - \underset{(1.1)}{21.2}\left(\frac{PCD}{PNS}\right)_t - \underset{(1.5)}{0.66}rr_t + \underset{(0.7)}{0.07}KCD_{t-1} - \underset{(0.6)}{0.35}UN_t - \underset{(6.3)}{3.43}ASTRIKE + \underset{(7.8)}{0.73}NCD_{t-1} + \underset{(1.4)}{0.28}UN_{t-1} + \underset{(4.4)}{0.45}FMOOD_t$$

$R^2 = .9955$; $SE = 1.31203$; $DW = 2.1129$; $\hat{\rho} = -\underset{(2.7)}{0.32}$

Zero-Effect

$$(19)\quad CD_t = \underset{(0.4)}{8.59} + \underset{(3.0)}{0.13}YD1_t - \underset{(0.4)}{0.04}\overline{YD1}^e_t - \underset{(0.8)}{17.08}\left(\frac{PCD}{PNS}\right) - \underset{(0.8)}{0.44}rr_t + \underset{(1.0)}{0.12}KCD_{t-1} + \underset{(0.1)}{0.07}UN_t - \underset{(6.3)}{3.32}ASTRIKE + \underset{(6.3)}{0.63}NCD_{t-1} - \underset{(0.2)}{0.14}UN_{t-1} + \underset{(4.6)}{0.48}FMOOD_t$$

$R^2 = .9957$; $SE = 1.28066$; $DW = 2.09$; $\hat{\rho} = -\underset{(2.1)}{0.24}$

An examination of the equations and their residuals indicates again that the zero-effect view dominates the full-effect view, although the difference is not nearly so striking as in the case of consumption. This is due to the fact that the coefficients on the income terms are quite small, so that adjusting income by a relatively small amount will make very little difference in the predicted value of spending on durables. The summary statistics are all better for equation (19) than for (18) and the residuals are smaller. However, it will be noted that the mean error over the first two years, as well as over the entire period, is uniformly positive under either view. This suggests that there is some other (unspecified) explanatory variable at work causing durables expenditures to be higher than expected. Therefore, the best that can be said about the surcharge is that it had no effect on durables spending. Evidently, the temporary drop in income *was* reflected in financial saving (see Table 5).

The results reported here for durables are really not very different from Okun's. He does not use total durables equations nor show us summary results for the total, but rather breaks it up into "autos" and "other durables." While he claims that his results from the other durables equation confirm the full-effect hypothesis, a closer examination of his own residuals and calculated *RMSE*s suggests that an hypothesis of 50 percent effectiveness would be closer to the mark. As for autos, he essentially ignores these equations, because the residuals imply that the surcharge led to significantly ***higher*** spending on cars, a finding which is completely at odds with expectations. Had he combined his results for both categories of durables, Okun would have been forced to the conclusion

TABLE 5—SAVINGS RATE

$$(\%) = 100 \times \frac{\text{personal saving, } NIA \text{ basis}}{\text{measured disposal income}}$$

Quarter	1967	1968	1969	1970	1971
I	7.4	7.2	5.3	6.9	8.2
II	7.0	7.6	5.3	8.1	8.6
III	7.5	6.0	6.6	8.3	8.1
IV	7.8	6.2	6.8	8.4	7.8

Source: Surv. Curr. Bus., July 1972.

that the surtax had no appreciable impact on this category of expenditures. Thus, his assertion that the surtax represented a successful application of flexible fiscal policy ultimately must hinge on his results for nondurables and services, which appear to differ markedly from my findings. It is to this discrepancy that I now turn.

C. *Why the Contradictory Results?*

1. In the first place, Okun's conclusion that the surcharge worked never was consistent with the incontrovertible fact that the saving rate took a sharp dip in late 1968 and throughout 1969 (see Table 5). It seems difficult to explain this other than by concluding that consumers maintained their spending patterns in the face of the surtax and financed this behavior by lowering their rate of saving.

2. Others have already questioned the validity of the conclusions Okun draws from his own analysis. Okun's assertion that the surcharge was 63 to 88 percent effective was obtained by imposing certain arbitrary constraints on his results. A more judicious interpretation of his findings would place the degree of effectiveness at something less than 50 percent.[14] Taken along with Okun's own statement that only one-half of the drop in disposable income was expected to be reflected in consumer spending (1971, p. 194), we are left with a surcharge impact of at best one-fourth that expected from a permanent change in taxes! Thus, Eisner's hypothesis of an *MPC* of .1 out of temporary income tax charges may not be as "wide of the mark" as Blinder and Solow maintain (p. 109).

3. My results are not strictly comparable to Okun's because the "consumption" variable includes depreciation and the implicit return on durables, along with expenditures on nondurables and services. Okun's equations, on the other hand, disaggregate consumer expenditures along the lines of the National Income Accounts. Therefore, in order to achieve stricter comparability the experiment was performed with an equation (using the same basic specification) in which nondurables and services was the dependent variable. This test (reported in my dissertation) merely confirmed the earlier finding that the surtax was ineffective in curbing consumption, regardless of whether or not it is defined to include the "use value" of durables. Such a result was to be expected since expenditures on nondurables and services constitute roughly 80 percent of the total value of consumption, while the use value on durables is essentially a trend which grows with the stock of real assets.

4. A possible source of bias in Okun's results may lie in the fact that the models which he used for simulating both zero- and full-effect views were all *estimated* using the full-effect concept of disposable income, (i.e., actual measured income). Thus, it is possible that his conclusions are thereby biased toward the full-effect view.[15] This question was explored by fitting my full-effect equation for consumption (16) to the data, but then performing the dynamic simulation using the zero-effect definition of income (which is how Okun tested the zero-effect view). Conversely, the zero-effect equation was simulated using full-effect income.

The results of this experiment (see the author for the details) pointed even more strongly to the superiority of the zero-effect view of the surcharge. The *RMSE* was considerably smaller for the zero-effect hypothesis (1.369) over the surcharge period than for the full-effect view (5.087).[16] Thus, when equations are estimated with one income concept and used to test the hypothesis concerning the other concept, the predicted values are not biased toward the income concept used in the estimation process. In this case, mixing the two definitions of income serves only to corroborate further the view that the surcharge had no appreciable impact on consumption. On a priori grounds, of course, it is preferable to test the alternative hypotheses by simulating equations which have been fitted using the corresponding definitions of income, as was done here.

---

[14] On this point, see Blinder and Robert Solow, William Branson, and comments by Eisner in Okun (1971).

[15] Okun (1971, p. 184) admits the possibility of such bias in his results, but does not pursue the matter.

[16] Such a result could have been anticipated from an

5. Two other possible explanations of the contradictory results were also explored, and the conclusions will be summarized here (see the author for further details). In the four forecasting models used by Okun, the independent variables consisted of some combination of income, lagged income, and lagged expenditure, plus assorted other *ad hoc* variables. None of these equations uses a real rate of interest, a variable which is grounded in theory and which does appear in my equations with considerable explanatory power. It therefore seemed appropriate to ask whether it was this interest rate variable which was making the zero-effect view come out better. However, when the experiment was performed with equations which excluded the interaction term, the basic finding of ineffectiveness of the surtax was not changed.

Secondly, a number of Okun's equations suffer from the problem of autocorrelated errors. Furthermore, it is likely that the autocorrelation in most cases is actually worse than it appears because of the presence of lagged dependent variables in many of the equations.[17] Since it has been found that estimated coefficients often change significantly when using ordinary least squares without the Cochrane-Orcutt transformation, I performed the experiment with an equation which did not employ this correction procedure. Again, however, the zero-effect view yielded smaller residuals over the surcharge period than did the full-effect view.

6. Having examined individually three possible sources of the discrepancy between Okun's and my results and failed to come up with a satisfactory explanation, the mystery was finally resolved by combining them. Thus, using an equation estimated by ordinary least squares, with the full-effect definition of income, and without an interest rate term, the equation was simulated over the fourteen-quarter period, 1968-III to 1971-IV, under the two alternative definitions of income. This is as close to Okun's procedure that I can get without completely changing my equations. The simulations now, of course, are static rather than dynamic. The results are shown below.

Full-Effect

$$CON_t = \underset{(0.02)}{-\,0.048} + \underset{(10.1)}{0.47\,YD_t} + \underset{(9.0)}{0.44\,\overline{YD}_t^{\,e}} \tag{20}$$

$R^2 = .9984$; $SE = 2.8737$; $DW = .3351$; (Ordinary Least Squares)

Comparing the error measures in Tables 6 and 7, except for the first four quarters following enactment of the surtax, the full-effect view outperforms the zero-effect view, in the sense that the average error is smaller. The period 1968-III to 1970-II is almost identical to Okun's simulation interval, and the *RMSE* is smaller for the

---

examination of the estimated coefficients on current income in equations (16) and (17).

[17] In general, the Durbin-Watson test is not applicable in situations where lagged values of the dependent variable are among the explanatory variables, and is biased toward two in such a case. On the other hand, presence of other exogenous variables in the equation tends to vitiate this difficulty somewhat. See Johnston, pp. 309–12, for a discussion of this matter.

TABLE 6—STATIC SIMULATION OF *Full-Effect* VIEW USING EQUATION (20)

(*CON* = Dependent Variable)

Period	*MAE*[a]	*RMSE*[b]	*ME*[c]
68-III to 69-II	2.175	2.559	2.175
68-III to 70-II	1.772	2.062	1.480
70-III to 71-IV	4.136	4.830	−4.136
68-III to 71-IV	2.786	3.536	−0.9272

[a,b,c] See Table 2.

TABLE 7—STATIC SIMULATION OF *Zero-Effect* VIEW USING *Full-Effect* EQUATION (20)

(*CON* = Dependent Variable)

Period	*MAE*[a]	*RMSE*[b]	*ME*[c]
68-III to 69-II	1.540	1.639	−1.540
68-III to 70-II	2.148	2.398	−2.148
70-III to 71-IV	5.352	5.831	−5.353
68-III to 71-IV	3.522	4.243	−3.522

[a,b,c] See Table 2.

full-effect view, although not by a very considerable margin.[18] Notice also that this is the only case where the mean error for the eight quarters, 1968-III to 1970-II, is positive on the full-effect view and negative on the zero-effect view. This means that in general, actual consumption fell between that predicted by the two extremes, rather than outside this interval as had been found earlier.

It thus becomes clear how these three defects of formulation and estimation procedure, when taken all together, would lead one to an erroneous conclusion as to which of the two views of the surcharge was the correct one on the basis of error-term analysis. We may speculate that had Okun used properly specified equations which included a real interest rate variable, had he made the appropriate correction for autocorrelated errors, and had he simulated the two hypotheses with equations estimated using the corresponding definition of income, he could not have escaped the conclusion that the 1968 income tax surcharge did not effectively reduce consumer expenditures.[19]

V. Concluding Comments

Arthur Okun's favorable assessment of the effectiveness of the 1968 temporary income tax surcharge puzzled a number of economists because it seemed to contradict what generally accepted consumer theory would predict for such a measure. His conclusion was doubly perplexing in view of the sharp drop in the savings rate following enactment of the tax, which suggested that most of the adjustment had occurred in financial savings flows. However, a closer and more balanced reading of Okun's findings indicates that his conclusion represents a somewhat overoptimistic interpretation of his own results, and that the actual impact on consumer expenditures was at best about one-fourth of what could be expected from a permanent tax change. And yet, the belief has persisted in many circles that flexible changes in income taxes are a useful and effective instrument of stabilization policy.

---

[18] A truly impartial analysis of the two simulations with equation (20) would actually suggest the hypothesis of 50 percent effectiveness as the most appropriate.

[19] Two further experiments were performed to test the validity of the results reported here. First, the basic equation was estimated using the "correct" procedures over the period 1955-I through 1967-IV, so as to exclude the surcharge period. Dynamic simulations were then run for the full- and zero-effect views over the period following enactment of the surtax. The *RMSE* for the zero-effect view (1.0) continued to be markedly smaller than that for the full-effect view (3.317) over the period 1968-III to 1970-II, and in fact the margin of superiority of the zero-effect view was considerably improved. It might be mentioned that three of Okun's four sets of equations were estimated over observations which included portions of the surcharge period. Secondly, the zero-effect/full-effect hypotheses were tested against the experience of the 1964 tax cut, which *permanently* lowered taxes on business and personal income. Using the basic consumption equation (11) and data on the amount of the tax cut reported in Okun (1968), the full-effect view was found to be superior to the zero-effect view. This result is consistent with what the permanent income theory would predict for a permanent tax change, and it confirms the earlier findings of Ando and Brown, and Okun (1968) that the tax cut was very successful in stimulating consumer spending.

This paper has attempted to lay this myth to rest once and for all by challenging Okun on his own ground of simulating consumption functions over the surcharge period using two extreme definitions of disposable income, one including and one excluding the tax. Equations for consumption and for expenditures on durables—both based on firm theoretical foundations—were estimated and found to have very small coefficients on current income, as well as a low adjustment coefficient on permanent income. This suggested a small short-run impact on consumption of any income tax change—permanent or temporary. Moreover, dynamic error-term analysis demonstrated unequivocally the superiority of the view that the surcharge had no appreciable impact on consumer expenditures. These results were most convincing in the case of consumption, as well as the somewhat more narrow category of consumption of nondurables and services—the very equations which in Okun's analysis had appeared to point most strongly to the superiority of the full-effect view.

Further, the causes of this inconsistency between my findings on the impact of the surcharge on consumption and Okun's were explored. It was found that the combination of failure to adjust for autocorrelated errors, the absence of a real rate of interest variable in the consumption equation, and the erroneous use of an equation estimated with the full-effect definition of income to simulate the zero-effect view would produce the counterintuitive conclusions obtained by Okun. The results described here do not suffer from these deficiencies and are consistent with what theory predicts in the case of a temporary income tax change.

Thus, the evidence of the 1968 surcharge period, far from confirming the general efficacy of temporary changes in income taxes, leads me to be pessimistic regarding the use of flexible tax policy to regulate aggregate demand. It appears that in order to have a significant effect on consumption expenditures, an income tax change must be permanent.

REFERENCES

A. Ando and **E. C. Brown**, "Personal Income Taxes and Consumption Following the 1964 Tax Reductions," in A. Ando et al., eds., *Studies in Economic Stabilization,* Washington 1968.

——— and **F. Modigliani**, "The 'Life Cycle' Hypothesis of Saving: Aggregate Implications and Tests," *Amer. Econ. Rev.,* Mar. 1963, *53,* 55–84; reprinted in R. A. Gordon and L. R. Klein, eds., *Readings in Business Cycles,* Homewood 1965.

A. S. Blinder, "Can Income Tax Increases Be Inflationary? An Expository Note," *Nat. Tax J.,* June 1973, *26,* 295–301.

——— and **R. M. Solow**, "Analytical Foundations of Fiscal Policy," in *The Economics of Public Finance,* Washington 1974.

W. H. Branson, "The Use of Variable Tax Rates for Stabilization Purposes," in R. A. Musgrave, ed., *Broad-Based Taxes: New Options and Sources,* Baltimore 1973.

——— and **A. K. Klevorick**, "Money Illusion and the Aggregate Consumption Function," *Amer. Econ. Rev.,* Dec. 1969, *59,* 832–49.

R. Eisner, "Fiscal and Monetary Policy Reconsidered," *Amer. Econ. Rev.,* Dec. 1969, *59,* 897–905.

———, "What Went Wrong?," *J. Polit. Econ.,* May/June 1971, *79,* 629–41.

R. C. Fair, "The Estimation of Simultaneous Equation Models with Lagged Endogenous Variables and First-Order Serially Correlated Errors," *Econometrica,* May 1970, *38,* 507–16.

———, *A Short-Run Forecasting Model of the United States Economy,* Lexington 1971.

M. Friedman, *A Theory of the Consumption Function,* Princeton 1957.

R. J. Gordon, "Inflation in Recession and

Recovery," *Brookings Papers*, Washington 1971, *1*, 105–66.

S. H. Hymans, "Consumer Durable Spending: Explanation and Predictions," *Brookings Papers*, Washington 1970, *2*, 173–206.

J. Johnston, *Econometric Methods*, New York 1972.

F. T. Juster and P. Wachtel, (1972a) "Anticipatory and Objective Models of Durable Goods Demand," *Amer. Econ. Rev.*, Sept. 1972, *62*, 564–78.

——— and ———, (1972b) "Inflation and the Consumer," *Brookings Papers*, Washington 1972, *1*, 71–121.

F. Modigliani and R. E. Brumberg, "Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data," in K. K. Kurihara, ed., *Post Keynesian Economics*, New Brunswick 1954; reprinted in H. R. Williams and J. D. Huffnagle, eds., *Macroeconomic Theory: Selected Readings*, New York 1969.

M. Nerlove, "The Dynamics of Supply: Estimation of Farmers' Response to Price," *Johns Hopkins University Studies in Historical and Political Science*, Series LXXVI, No. 2, Baltimore 1958.

A. M. Okun, "Measuring the Impact of the 1964 Tax Reduction," in W. W. Heller, ed., *Perspectives on Economic Growth*, New York 1968.

———, "The Personal Tax Surcharge and Consumer Demand, 1968–70," *Brookings Papers*, Washington 1971, *1*, 167–212.

W. Poole, "The Role of Interest Rates and Inflation in the Consumption Function," *Brookings Papers*, Washington 1972, *1*, 211–20.

W. L. Springer, "Windfalls, Temporary Income Taxes, and Consumption Behavior," unpublished doctoral dissertation, Princeton Univ. 1974.

U.S. Office of Business Economics, *Surv. Curr. Bus.*, July 1972, *52*.

Linear Taxation of Wealth and Wages for Intragenerational Lifetime Justice: Some Steady-State Cases

By J. A. Ordover and E. S. Phelps*

Every paper on the intrageneration redistribution of utilities rightfully makes reference to James Mirrlees. His celebrated paper is the first to study a problem in optimum redistribution without the magic wand of the lump-sum tax. What has proved seminal in that paper is its attractive representation of generational heterogeneity: all worker-consumers can be placed along a one-dimensional continuum with respect to their different efficiencies at producing. Otherwise they are all alike, are equally efficient at consuming, and have the same egoistic utility function that each individual maximizes.

The setting for the analysis is a stark one-period competitive-market economy of bread and leisure. Total output of bread is a constant-marginal-returns function of the aggregate efficiency-weighted man-hours worked—there is no capital. The wage paid to an individual is equal to his product, the efficiency-weighted hours he elects to work. The tax he pays is some function (alike for all individuals) of his wage income. The aggregate tax revenue thus collected is (after subtractions) disbursed to all individuals in the form of a demogrant or lump-sum credit equal for all individuals. Mirrlees's redistributional problem, then, is to find a tax-net-of-demogrant schedule against wage income that maximizes the additive Bentham-Bergson-Fleming social-welfare functional, $\int u(n)^{-\beta} dn$.

It has been the source of some astonishment that no problem in optimum redistribution by non-lump-sum means was even stated, let alone analyzed, until 1971. It is also surprising that the problem that came eventually to be formulated deals with an ornery "detail"—the shape or progressivity of a particular tax schedule. It might have been expected that the economics of redistribution would begin with the analysis of a problem both easier and perhaps more important, to wit: the optimum *mix* of *proportionate* taxes with which to finance redistributive transfers. The marginal utility of departures from linearity in the tax functions could always be left for later attention. Of course, to study such a problem one needs to endow the model economy with more than one factor of production (or more than one consumer good).

Our paper investigates the optimal mix of taxes on two factors of production, capital and (efficiency-weighted) labor. More precisely, we analyze the optimum proportionate taxation of wages and interest, or wages and wealth, with account taken of the public debt. The vehicle for analysis is a blend of Mirrlees's theory of efficiency-weighted labor supply and the life-cycle theory of the supply of wealth found, for example, in the growth model constructed by Peter Diamond.[1]

---

* New York University and Columbia University, respectively. The present paper is an abridged version of a much more extensive working paper cited in the bibliography. Ordover gratefully acknowledges partial support from the National Science Foundation grant to New York University.

The particular optimization problem studied here possesses two special features. The social welfare function we employ expresses the conception of economic justice championed by John Rawls: As a chain is no stronger than its weakest link, social welfare is only as great as the utility of the least well-off, the person or persons having minimum utility. Thus the redistributional optimum is "maximin."

The other feature is a restriction on the dynamics of capital and public debt. The basic notion is also found in a related study of interest taxation by Ordover, in which society is constrained to maintain at some predesignated level the ratio of capital to (geometrically progressing) population. We may not know to what quantity that ratio ought "in justice" to be equal. But whatever that quantity and whenever it is to be attained, society will want then to have that mix of taxes (and resulting demogrant) which is economically just *intragenerationally*, subject to the mandate that fiscal policy preserve this capital-population ratio for the use of future generations. As for the ratio of public debt to population, this too is to be maintained constant from generation to generation, but at a preoptimized rather than arbitrary level.

The present paper is more extensive. We first formulate the "general" (steady-state) problem in which both the public debt and the capital stock per worker are arbitrary constants. We then report our analysis of the optimal tax structure in the aforementioned problem and in the diagonally opposite problem—where the debt per worker is fixed arbitrarily and *capital* per worker is maintained intact at its optimal steady level. Lastly we cite some results obtained for the joint problem of the Golden Rule steady state in which both capital and debt per worker are maintained at their best steady levels.

The paper here also differs with regard to the channels of tax "distortion." As in the original Mirrlees paper, the disincentive effect of taxation falls on the quantity of hours worked (as well as the volume of private saving) rather than the quantity of investment in education, as in the "Polish view" adopted by Ordover and earlier Eytan Sheshinski. As a consequence the supply of efficiency-labor depends not only upon the after-tax wage rate paid to a unit of efficiency-work but also upon the after-tax rate of return to saving—the latter a factor emphasized by several general-equilibrium theorists from Hicks to Lucas.

Our emphasis is on the theory of optimal taxation under market capitalism, with its obvious agenda of questions: Should profits be taxed? Should they be taxed more heavily than wages? Might it be optimal to levy a proportionate wealth tax at a rate exceeding the average rate of profit on capital?[2]

The model has relevance to the theory of market socialism as well as to the welfare economics of capitalist taxation. Karl Marx, Oskar Lange, and some other socialist theorists explained that the returns to capital garnered by the state after payment of wages would be distributed to the people as an equal poll-subsidy/social-divided after subtraction of the state's expenditures on final goods plus any budgetary surplus appropriate to its plans for capital formation. They did not tell us whether a household might be allowed to accumulate wealth claims to future goods; nor did they say what rate of return—positive, zero, or negative—such abstract property, if allowed, ought to be paid for maximum social welfare. The present paper sheds some light on those unanswered questions.

---

[1] Koichi Hamada has studied a different problem in the same setting.

[2] The answers are that "it depends"—depends upon the quantities of capital and public debt per worker and upon the elasticities of factor supplies and demands. The novelty here is that the answers are "technical," being reasoned from an explicit ethical postulate, namely, the maximin criterion. (Use of this analytical approach does not imply indifference or obliviousness to institutions, considered as ends or as means.) In some other approaches, it almost seems as though there is intended to be some kind of justice between people and machines, an animism in which machines usually are held morally inferior to humans.

I. The Model

We first discuss production possibilities, next markets and accounting, then consumer behavior.

A. *Resources and the Production Set*

Each generation lives for two periods. Its members work, if at all, over the first period; they consume, if at all, at the end of each of these two periods. The size of the working age population, N, grows exogenously at the geometric rate $n \geq 0$:

$$(1) \qquad N_{-1} = (1+n)^{-1}N$$

Each per worker variable used below is a certain aggregate divided by current N. We shall be constraining the economy to maintain a steady state in which by definition each such per worker variable is equal from generation to generation.[3]

The per worker quantity of effective labor supplied by the current workforce, denoted by l, is measured in standardized units of "efficiency man-hours." It is an endogenous variable, a function of the after-tax prices and transfers. The efficiency of each person's hours worked is measured by a parameter m that ranges from zero to some $M>0$. The history of any individual of type m is denoted by

$$(2) \qquad x^m = (x_0^m, x_1^m, x_2^m, x_3^m), \qquad m \in [0, M]$$
$$x_0^m + x_3^m = 1, \quad x_j^m \geq 0 \quad \text{all } j$$

where x_0^m is first period leisure, x_i^m is consumption at the end of age i ($i=1, 2$), and x_3^m is "effort" or "worktime" in the first period of life. Worktime is measured in *natural units*, namely as a fraction of the duration of the period.

The basic postulate here is that efficiency differences among individuals are purely labor augmenting; that is, the marginal rate of substitution in production between a man-hour of type m_1 and a man-hour of type m_2 is "constant," given simply by m_1/m_2, independent of the quantities of capital and labor types working alongside. It follows that there exists a labor aggregate which adds up the various man-hours worked when converted to efficiency units. To obtain from the quantity of effort x_3^m, the amount of labor service in efficiency units that it supplies, say l^m, we need only multiply the former by m.[4] Thus we are standardizing on the productivity of the $m=1$ types. Now let the proportion of the population of any and every generation having an efficiency *m or less* be given by a nondecreasing cumulative-distribution function, Φ, with the properties

$$(3) \qquad 0 < \Phi(0) < 1, \ \Phi(M) = 1,$$
$$\Phi(a) \leq \Phi(b) \qquad \text{if } 0 \leq a < b \leq M$$

Then per worker effective labor supplied is[5]

$$(4) \qquad l \equiv \Sigma l^m = \Sigma m x_3^m$$

and the other per worker quantities are

$$(5) \qquad x_i \equiv \Sigma x_i^m, \qquad i = 0, 1, 2, 3$$

We posit a standard one-sector neoclassical net aggregate output function, F. In any steady state with constant k, a certain portion nk of per worker net output,

---

[3] Note that N is the potential number of workers in the market sector, an upper bound on the number who choose to work.

[4] There is no loss of generality in multiplying by m instead of a monotone function of m because the frequency function of m can be adjusted as needed.

[5] For convenience of notation we replace the integral sign by the summation sign.

$F(k, l)$, must be allocated to net investment to keep k constant. Per worker government consumption, $\gamma \geq 0$, is likewise constant. The remaining fraction of net output is available for per worker consumption c.[6] The latter is distributed in a steady state between the per worker consumption of current workers and retired workers thus:[7]

$$c = x_1 + x_2(1+n)^{-1} \tag{6}$$

Hence steady-state growth with constant k entails the *capital constraint*

$$F(k, l) - x_1 - x_2(1+n)^{-1} = \gamma + nk \equiv \text{constant} \geq 0 \tag{7}$$

B. *Market Organization, Prices, and Social Accounting*

We turn now to the demand side of the model and to the market aspects.[8] There are three goods and we may think of two perfect markets: a labor market where labor (homogeneous after conversion to efficiency units) is exchanged for present consumables, and a capital market where savings (unexercised claims to present goods) are traded for claims to future goods.

So far as proportionate taxation of households' market demands and supplies are concerned, there is no loss of generality in our assuming that expenditures on the commodities (x_1 and x_2) are untaxed. Hence let us consider a proportionate tax rate on wage income, denoted τ_w, and a proportionate tax rate on profits (or interest) from saving, τ_r. (We can also entertain the possibility of a proportionate tax on private wealth.) In the "natural" interpretation, then, our two markets *and* taxation determine *four* relative prices of the two factors: The before-tax real wage rate w, the after-tax real wage rate ω, the rate of return before tax on real private savings r, and the after-tax rate of return ρ. The term "real" denotes deflation by the "money" price of presently consumable commodities, and the term "wage rate" refers to the rate for a standard individual, the real wage per efficiency unit of labor.

We postulate the universal lifetime income guarantee, or demogrant, which comes in two installments, β_1 and β_2. For the moment we think of β_1 and β_2 as each nonnegative. Accordingly there is an obvious lifetime budget constraint for households of *any* type m which, when aggregated, yields the relation

$$(1+\rho)^{-1}(x_2 - \beta_2) = \omega l - (x_1 - \beta_1) \tag{8}$$

Note that the right-hand side of (8) equals the per worker purchases of wealth by the present working-age generation. But aggregate wealth in a steady state must grow geometrically at rate n. Therefore, if d denotes the per worker stock of public debt existing at the beginning of the period, and so $k+d$ is the per worker quantity of wealth, then the workers must purchase $(1+n)(k+d)$ of wealth in order that per worker wealth be unchanged next period. Hence the *private wealth constraint:*

$$(1+\rho)^{-1}(x_2 - \beta_2) = (1+n)(k+d) = \text{constant} > 0 \tag{9}$$

Equations (8) and (9) imply the *social budget constraint:*

$$x_1 + \frac{x_2}{1+n} = \beta_1 + \frac{\beta_2}{1+n} + \omega l + \rho(k+d) - n(k+d) \tag{10}$$

---

[6] Obviously the requirements of growth and public expenditure may be too great to admit a solution with nonnegative consumption and efficiency-weighted leisure.

[7] In a state of steady growth at rate n, the (future) second-period consumption x_2 of the present generation of first-period people is $(1+n)$ times the second-period consumption of people now in their second period.

[8] For an extensive discussion of these matters, the interested reader may consult the authors' working paper.

This states that, with regard to aggregates, consumption equals disposable income *less* private saving. Let us also observe that, expenditure taxes nil, national output equals national income:

$$(11) \qquad rk + wl = F(k, l)$$

Hence, from (7)

$$(12) \quad x_1 + \frac{x_2}{1+n} = rk + wl - \gamma - nk$$

Subtracting (10) from (12) we obtain the government's income statement

$$(13) \quad (r-\rho)k + (w-\omega)l - \rho d = \beta_1 + \frac{\beta_2}{1+n} + \gamma - nd$$

The left-hand side is net tax revenue—net, that is, of after-tax interest on the public debt; this must cover public benefits and expenditures *less* the budgetary deficit. In (13), the deficit is constrained to maintain the given debt-worker ratio d_0, however determined—whether predetermined by actual history or hypothetically optimized. Hence (13) expresses a *public debt constraint* that, using (11), we may state more explicitly and analogously to (7) in the form

$$(14) \quad F(k,l) - \rho(k+d) - \omega l - \beta_1 - \frac{\beta_2}{1+n} = \gamma - nd_0 = \text{constant}$$

The capital constraint in (7) prescribes a certain amount of national saving while the debt constraint prescribes how much of this shall be public saving, $-nd_0$, in lieu of private saving.[9]

C. *Consumer Behavior*

Regarding consumers, we suppose that households of any type m choose $x^m = (x_1^m, x_2^m, x_3^m)$ to maximize utility $u(x^m)$, which is independent of m, subject to their budget constraint—like (8), expressed here in terms of present value or discounted prices, $q^m = (q_1, q_2, q_3^m)$. The latter are defined thus:

$$q_1 \equiv 1, \quad q_2 \equiv (1+\rho)^{-1}, \quad q_3^m \equiv m\omega, \quad q_3 \equiv$$

In these terms the maximization problem is

$$(15) \quad V^m(q^m; \beta_1, \beta_2) = \max_{\{x^m\}} \left\{ u(x^m) - \alpha^m \left[\sum_1^2 q_i^m (x_i^m - \beta_i) - q_3^m x_3^m \right] \right\}$$

V^m is the indirect utility function, homogeneous of degree zero in the prices, with derivatives

$$(16) \qquad \begin{aligned} \partial V^m / \partial q_3^m &= \alpha^m x_3^m \\ \partial V^m / \partial q_2^m &= -\alpha^m x_2^m \\ \partial V^m / \partial \beta_i &= \alpha^m q_i \qquad i = 1, 2 \end{aligned}$$

where the Lagrange multiplier $\alpha^m > 0$ is the marginal utility of (present) money.

From the households' utility maximizations we can obtain the individual demand functions for x_1^m, x_2^m, and the individual supply functions for saving $\sigma^m \equiv q_2(x_2^m - \beta_2)$ and effective labor l^m. Aggregating these demands and "factor" supplies, we obtain the per worker market demands for x_1 and x_2, and the market supplies of labor l and wealth $\sigma \equiv q_2(x_2 - \beta_2)$. These market demands and supplies are each a function, given β_2, of the triplet $(1, q_2, q_3, \beta_1)$. These market functions must obey the lifetime budget constraint in (8), which we write here

$$(8') \quad q_2(x_2 - \beta_2) = q_3 l - x_1 + \beta_1 \equiv \sigma$$

Let σ_i denote $\partial\sigma/\partial q_i$, and similarly for l_i, x_{1i}, and x_{2i}. Then differentiation of this budget equation yields

---

[9] The constraint on public saving in (14) could be replaced by the constraint on *current* private *all-generation* saving in (10) or, as well, by the constraint on after-interest private wealth in (9). But the overlife constraint on the present generation only is no substitute for these social constraints.

$$
\begin{aligned}
(17)\quad x_2 - \beta_2 + q_2 x_{22} &= 0 + q_3 l_2 - x_{12} \equiv \sigma_2 \\
q_2 x_{23} &= l + q_3 l_3 - x_{13} \equiv \sigma_3 \\
q_2 x_{2\beta} &= 1 + q_3 l_\beta - x_{1\beta} \equiv \sigma_\beta
\end{aligned}
$$

To narrow the analysis that follows, we wish to place some natural restrictions on the functions x_1, x_2, l, and σ. Thus, we shall suppose that all goods are strictly noninferior, subject to some inessential qualifications at boundaries, so that

$$
\begin{aligned}
(18)\quad & x_{i\beta}>0,\ 0<q_i x_{i\beta}<1, \quad i=0, 1, 2;\ x_{3\beta}<0 \\
& l_\beta<0,\ \sigma_\beta>0 \qquad (\text{for } q_2<\infty,\ q_3>0)
\end{aligned}
$$

Concerning the wealth supply function we shall assume, *first*, that $\sigma(1, \infty, q_3, \beta_1)=0$: that is, no one will save at the confiscatory $\rho=-1$. (Since there may be public saving, the latter does not guarantee that $\rho=-1$ is inoptimal.) *Second*, a rise of the wage rate increases saving everywhere. Then, for all $q_2<\infty$ and $q_3>0$,[10]

$$
(19)\qquad \sigma_3(1, q_2, q_3, \beta_1) = q_2 x_{23} > 0
$$

Third, because the "marginal propensities" $q_i x^m_{i\beta}$ are between zero and one, by noninferiority, we can also argue that

$$
(20)\quad \sigma_2(1, q_2, q_3, \beta_1) = \begin{cases} <0 & \text{at } q_2 = \infty \\ >0 & \text{at } q_2 = 0 \end{cases}
$$

The argument is that σ increases with ρ in the neighborhood of $\rho=-1$, because there is no income effect to counter the positive substitution effect of larger ρ. Hence, σ "bends backward" at sufficiently large ρ (small q_2) in the (σ, ρ) plane and likewise in the (σ, q_2) plane.

Concerning the labor supply function we shall assume, *first*, that $l(1, q_2, 0, \beta_1)=0$: no one will work at $\omega=0$. *Second*, we specify that

$$
(21)\quad l_2(1, q_2, q_3, \beta_1) = \begin{cases} \lesseqgtr 0 & \text{at } q_2 = \infty \\ >0 & \text{at } q_2 = 0 \end{cases}
$$

As $q_2\to\infty$ the income effect of higher q_2 becomes negligible, so l_2 has the sign of the cross-substitution effect, positive or negative. If negative then there must hold a *Hicks-Lucas* effect $l_2<0$ at sufficiently large q_2—that is, labor supply must be increasing in ρ around $\rho=-1$. As $q_2\to 0$, $x_2\to\infty$, so the income effect of a rise in q_2 is presumed to be decisive, decreasing leisure and increasing labor supply, whatever the sign of the cross-substitution effect. We shall refer to $l_2>0$ as the *anti-Hicks-Lucas effect.*

Finally, we have

$$
(22)\quad l_3(1, q_2, q_3, \beta_1) = \begin{cases} >0 & \text{at } q_3 = 0 \\ =0 & \text{at } q_3 = \infty \end{cases}
$$

The labor supply function is increasing at $q_3=0$, where $l(\cdot)=0$, and it may bend backward at sufficiently large q_3.

II. The Analysis

Our problem is the selection of tax rates on wages and interest (or wealth) so as to maximize over feasible steady states the minimum level of utility, $\min_m V^m(\cdot)$. This bottom utility must equal $V^0(\cdot)$, for no one of any type m could do worse than the $m=0$ types. As is widely done in the optimal taxation literature, we actually maximize $V^0(\cdot)$ with respect to the after-tax factor rewards (ρ, ω) or equivalently (q_2, q_3), as if the optimization were for a socialist state. This is merely a calculating device. To each pair (ρ, ω) there corresponds a unique tax-rate pair (τ_r, τ_w) and the latter is often readily solved for.

This maximization is subject in the "general" case to any *two* of the three constraints—(7) on capital, (10) on private wealth supply, (14) on public debt. In terms of the first two of these constraints, the general problem can be expressed for any given β_2 as

$$
(23)\qquad \underset{\{q_2, q_3, \beta_1\}}{\text{maximize}}\ V^0(q_2, q_3, \beta_1; \beta_2)
$$

---

[10] An interpretation of this inequality, though not a necessary implication, is that leisure and second-period consumption are everywhere net substitutes.

subject to

$$(23') \quad nk = F(k, l) - (q_3 l + \beta_1 - \sigma) - (q_2^{-1}\sigma + \beta_2)(1 + n)^{-1} - \gamma \equiv J(q_2, q_3, \beta_1; k, \beta_2)$$

$$(23'') \quad (1 + n)(k + d) = q_2(x_2 - \beta_2) \equiv \sigma(q_2, q_3, \beta_1; \beta_2)$$

In (23′) we have obviously substituted for x_1 from the lifetime budget constraint (8′) in terms of l and σ. Similarly, we represent x_2 by the same function of $(q_2, q_3, \beta_1, \beta_2)$, namely $q_2^{-1}\sigma + \beta_2$, as gives the *ex ante* x_2 of the current workers on the ground that the prices established must be the same over generations in any steady state.

The first-order conditions for a maximum in (23) involve the first derivatives V_2^0, V_3^0, and V_β^0 of the indirect utility function V^0. We have

$$(24) \quad \overset{0}{V}_2 = -(\overset{0}{x}_2 - \beta_2)\overset{0}{V}_\beta, \quad \overset{0}{V}_3 = \overset{0}{l}\overset{0}{V}_\beta = 0, \quad \overset{0}{V}_\beta > 0$$

Here, V_β^0 is the marginal utility of β_1 to the poor; V_3^0 is zero because the $m=0$ type cannot earn wages.

The sign of V_2^0 is indeterminate until β_2 is specified. We shall confine the analysis reported here to the following special case: The parameter β_2 has been fixed such that, *at the maximum in* (23),

$$(25) \quad \beta_2 = \overset{0}{x}_2 \geq 0$$

To motivate (25) it might be argued that the government would want β_2 large enough to lift from the poor the burden of providing for their old age out of their β_1 yet not so large that the poor are induced to borrow against future β_2 for additional first-period consumption.

Consequently $V_2^0 = V_3^0 = 0$ in the neighborhood of the maximum. A change of q_2 or q_3 in this neighborhood will therefore increase $V(\cdot)$ only by increasing β_1 via one of the constraints. Hence the pair (q_2, q_3) which maximizes $V(\cdot)$ in (23) must be simultaneously maximizing β_1.

The maximization of β_1 may be formulated as a problem in what might be termed "stationary programming." Noting that $J_\beta < 0$, let us assign β_1 to meet the capital constraint in (23′). Then choose (q_2, q_3) to maximize β_1 so determined, subject to the wealth constraint in (23″). (Because $\sigma_3 > 0$ it is also natural to think of q_3 as tied to the wealth constraint, thus making q_2 the variable free for decision.) The *maximum* value of β_1 depends only on the exogenous k and d (given β_2 and the other parameters) and is denoted $b(k, d)$:

$$(26) \quad b(k, d) = \max_{\{q_2, q_3\}} \beta_1$$

subject to (23′) and (23″).

From (26), (23′), and (23″) we then obtain the functional equation

$$(27) \quad 0 = \max_{\{q_2, q_3\}} \{J(q_2, q_3, b(k, d); k) - nk - \lambda[\sigma(q_2, q_3, b(k, d)) - (1+n)(k+d)]\}$$

In this maximization, $b(k, d)$ and the Lagrange multiplier $\lambda(k, d)$ are "constants," independent of (q_2, q_3), being functions only of the predetermined state variables (k, d) and the other parameters.

The interpretation of (27) is clear. To maximize β_1, (q_2, q_3) are chosen, subject to $\sigma = (1+n)(k+d)$, to make $J(\cdot)$ as large as possible. For since $J_\beta < 0$, the larger is $J(\cdot)$ at any β_1, the higher is the β_1 that keeps $J(\cdot) = nk$.

At the maximum β_1 the following first-order conditions hold:

$$(28a) \quad J_2 - \lambda\sigma_2 = 0, \qquad J_3 - \lambda\sigma_3 = 0$$

$$(28b) \quad J_\beta - \lambda\sigma_\beta \leq 0$$

The argument for (28b) is that if it did not hold there would be room for some free self-sustaining rise of β_1 which would increase $\lambda\sigma(\cdot)$ by so much as to permit (via q_2 and q_3) a net increase of $J(\cdot)$, hence still larger β_1; but such is impossible at the maximum β_1.

By equating to zero the total derivative with respect to d of the right-hand side of (27) we also obtain the "marginal worth" of an increase in the debt per worker:

$$b_d(k, d) = \frac{\lambda(1+n)}{-J_\beta + \lambda\sigma_\beta} \tag{29}$$

It follows from (28b) that λ and $b_d(k, d)$ are like-signed—positive, negative, or zero as the debt is too small, too large, or just right. It may be noted that λ and hence $b_d(k, d)$ have the sign of J_3 by virtue of (28a) and the specification that $\sigma_3 > 0$. If optimal $J_3 < 0$, for example, a larger debt, by entailing larger $\sigma(\cdot)$ and therefore larger q_3 at each q_2, would result in smaller $J(\cdot)$ at any given β_1 and thus imply a reduction of maximum β_1, that is, $b_d(k, d) < 0$.

A. *Optimal Taxes with Arbitrary Debt and Capital*

Space permits only fleeting attention here to the tax implications of the general problem as studied in our working paper. The succeeding sections on the more restrictive problems will serve to indicate the kinds of propositions developed and to convey the analytics employed.

For use in Sections IIA, B, and C, respectively, we define these *specific* (as distinct from *ad valorem*) tax rates:

$$\begin{aligned} \text{(30a)} \quad t_\lambda &= (1-\lambda)(1+\rho)^{-1} - (1+n)^{-1} \\ &= [n - \lambda(1+n) - \rho]q_2(1+n)^{-1} \\ \text{(30b)} \quad t_n &= (1+\rho)^{-1} - (1+n)^{-1} \\ &= (n-\rho)q_2(1+n)^{-1} \\ \text{(30c)} \quad t_r &= (1+\rho)^{-1} - (1+r)^{-1} \\ &= (r-\rho)q_2(1+r)^{-1}, \qquad r \equiv F_k \\ \text{(30d)} \quad t_w &= w - \omega, \quad \tau_w = t_w\omega^{-1} \quad w \equiv F_l \end{aligned}$$

Evidently t_n is a kind of "shadow" tax rate. It measures the wedge between ρ, the after-tax rate of return to saving, and n, the "natural" rate of interest—essentially, the marginal rate of transformation between x_1 and x_2 in the equation $J - nk = 0$. For $\lambda \neq 0$, however, t_λ is the analogous shadow tax rate. It appears if we write our maximand in (27) as

$$\text{(27}') \quad \underset{\{q_2, q_3\}}{\text{maximize}}\; F(k, l) - q_3 l + t_\lambda x_2 + \text{constant}$$

where again the functions $l(q_2, q_3, \beta_1)$ and $x_2(q_2, q_3, \beta_1)$ are to be evaluated at $\beta_1 = b(k, d)$. The derivative of the maximand with respect to efficiency-labor is $t_w = F_l - q_3$ and the derivative with respect to x_2 is t_λ.

The first-order conditions for a maximum in (28) can thus be expressed as:

$$\begin{aligned} \text{(31)} \quad l_2 t_w + x_{22} t_\lambda &= -(1-\lambda)(x_2 - \beta_2) \\ l_3 t_w + x_{23} t_\lambda &= l \\ l_\beta t_w + x_{2\beta} t_\lambda &\leq 1 \end{aligned}$$

While a few results, largely in the nature of logical possibilities, are derivable from (31), we proceed now to the more restrictive problem with preoptimized debt in Section B and with preoptimized capital in Section C.[11]

B. *Optimal Taxes with Optimized Debt*

In the problem here, public debt per worker is a free variable. The portion of investment comprised of *private* saving is unconstrained, subject only to *nonnegativity* of private wealth supply. In the special case defined in (25) therefore, where $V_2^0 = 0$, the optimal tax problem may be expressed as

---

[11] An amusing footnote to the general problem of exogenous k and d is its Cantabridgian theory of the functional distribution of *disposable* national income between interest, $\rho(k+d)$, and wages, ωl:

> *Cambridge Lemma*: Because all interest (after tax) is consumed while some wages (after tax) are saved, the functional distribution of income (after tax), $\{\rho(k+d), \omega l\}$ must favor wages just enough to insure that private saving equal $n(k+d)$.

Of course, it is the factor unit rewards, ω and ρ, which are of policy significance.

$$nk = \max_{\{q_2, q_3\}} J(q_2, q_3, b(k); k) \tag{32}$$

subject only to

$$0 \leq q_2(x_2 - \beta_2) \equiv \sigma(q_2, q_3, b(k)) \tag{32'}$$

where $b(k)$ is the maximized β_1.

At an interior solution, where the nonnegativity constraint is not binding, the first-order conditions are

$$J_2 = 0, \quad J_3 = 0; \quad J_\beta < 0 \tag{33}$$

A subsequent proposition will address the possibility of a corner solution at $\sigma(\cdot) = 0$.

The following properties of the solution can now be shown.

PROPOSITION B1: *Optimal* $\omega > 0$ *whether or not optimal* $\sigma(\cdot) > 0$.

If $\omega = 0$, $l = 0$ whence $J(\ ) < 0 < nk$ which is inadmissible.

PROPOSITION B2: *The corner solution at* $\sigma(\cdot) = 0$ *occurs if and only if optimal* $\rho = -1$.

First note that $\sigma(\rho, \omega, \beta) = 0$ for all ω and β if $\rho = -1$, that is, $q_2 = \infty$. Conversely, the existence of a solution at $\sigma(\cdot) = 0$ implies $\rho = -1$. For $\omega > 0$ by Proposition B1, $\beta_1 > 0$ for the solution to permit $0 < x_1^0 (= \beta_1)$, while $\sigma(\rho, \ \omega > 0, \ \beta_1 > 0) > 0$ for all $\rho > -1 (q_2 < \infty)$. So the optimal allocation will display property owning if optimal $\rho > -1$.

PROPOSITION B3: *Optimal* $l_3 > 0$ *whether the solution is at the corner or is interior.*

At an interior optimum, $J_3 = 0$. At the corner solution where $\sigma(\cdot) = 0$, $\rho = -1$ so that $\sigma_3 = 0$ for all ω. Hence $J_3 = 0$ at a corner solution as well. Recall now that

$$J(\cdot) = F(k, l) - x_1 - x_2(1 + n)^{-1} - \gamma$$

Since $F_l > 0$, $x_{13} > 0$, and $x_{23} \geq 0$, $l_3 \leq 0$ would imply $J_3 < 0$, a contradiction. The economics here is that at an allocation on some backward-bending segment of the labor-supply curve, where $l_3 < 0$, it would be possible to drop ω to a rising segment for a gain in $J(\cdot)$ at the initial β_1 and thus obtain some increase in β_1 without violating $J(\cdot) = nk$.

PROPOSITION B4: *Optimal* $\rho = -1$ *implies* $l_2 \geq 0$ *(no Hicks-Lucas effect) in that neighborhood. Hence the presence of a Hicks-Lucas effect of larger* ρ, *at least around* $\rho = -1$, *is sufficient to rule out a corner solution.*

Using the budget relation $x_1 = \omega l + \beta_1 - (1+\rho)^{-1}(x_2 - \beta_2)$ we write

$$\begin{aligned} J(\cdot) = {} & F(k, l) - \omega l - \beta_1 \\ & + (x_2 - \beta_2)(1 + \rho)^{-1} \\ & - x_2(1 + n)^{-1} - \gamma \end{aligned} \tag{34}$$

At the corner where $\rho = -1$ and thus $x_2 = \beta_2$ the first-order conditions for a constrained maximum are

$$\left.\frac{\partial J}{\partial \rho}\right|_{\rho=-1} = (F_l - \omega)\frac{\partial l}{\partial \rho} + t_n \frac{\partial x_2}{\partial \rho} \leq 0 \tag{35a}$$

$$\left.\frac{\partial J}{\partial \omega}\right|_{\rho=-1} = (F_l - \omega)\frac{\partial l}{\partial \omega} - l = 0 \tag{35b}$$

recalling that $x_{23} = 0$ at $\rho = -1$. Now $t_n \equiv (1+\rho)^{-1} - (1+n)^{-1} > 0$ at $\rho = -1$. And $t_w = F_l - \omega > 0$ since $l > 0$ for a solution and $l_3 > 0$ by Proposition B3. Also $\partial x_2/\partial \rho \geq 0 (x_{22} \leq 0)$ around $\rho = -1$. Hence (35a) implies $\partial l/\partial \rho \leq 0$ which is non-Hicks-Lucas. Actually, if $\partial x_2/\partial \rho > 0$, $\partial l/\partial \rho < 0$.

In the remainder of this section we shall be assuming that the solution is interior at some $\sigma(\cdot) > 0$. This means that the labor-supply function is Hicks-Lucas at least around $\rho = -1$. Then $\rho > -1$ and the first-order conditions in (33), expressed in terms of (q_2, q_3, β_1) and the tax rates (t_w, t_n), are

$$l_2 t_w + x_{22} t_n = -(x_2 - \beta_2) < 0 \tag{36a}$$

$$l_3 t_w + x_{23} t_n = l > 0 \tag{36b}$$

$$l_\beta t_w + x_{2\beta} t_n < 1 \tag{36c}$$

PROPOSITION B5: *Either t_w or t_n or both are positive.*

Since $l_3>0$ and $x_{23}>0$, $t_n\leq 0$ implies $t_w>0$ from (36b).

PROPOSITION B6: *If $l_2\geq 0$ at the optimum, then $t_n>0 (\rho<n)$.*

Assume the contrary, that $t_n\leq 0$. Then $t_w<0$ when $l_2\geq 0$ by (36a). But $t_w>0$ if $t_n<0$ by (36b)—Proposition B5. Hence $t_n>0$ if $l_2\geq 0$.

To the same end, we may solve (36a) and (36b) for t_n and t_w:

$$(37)\qquad t_n=\frac{ll_2+(x_2-\beta_2)l_3}{l_2x_{23}-l_3x_{22}}$$

$$t_w=\frac{-(x_2-\beta_2)x_{23}-lx_{22}}{l_2x_{23}-l_3x_{22}}$$

If $l_2\geq 0$, then the common denominator is positive and so is t_n. But t_w may apparently be of either sign.

The presence of anti-Hicks-Lucas effect $l_2\geq 0$ at the optimum rules out $\rho>n$ as we have just seen. One might expect also that it precludes a solution where σ is backward-bending: that is, $\partial\sigma/\partial\rho<0$ or $\sigma_2>0$. But we can prove only the conditional statement:

PROPOSITION B7: *Should the optimum be anti-Hicks-Lucas ($l_2\geq 0$), then $\sigma_2<0$ if optimal $t_w\geq 0$.* COROLLARY: *If $\sigma_2\geq 0$ at such optimum, then $t_w<0$.*

To prove, we note that

$$(38)\quad J_2=t_wl_2+\sigma_2-(1+n)^{-1}x_{22}=0$$

Then $t_w\geq 0$ and $x_{22}<0$ imply $\sigma_2<0$. The corollary is immediate and may be interpreted as follows: An increase of ρ within any backward-bending stretch of σ must, apart from its effect on labor supply, reduce $J(\cdot)$ because it must reduce $t_n(x_2-\beta_2)$—certainly as long as $t_n>0$, as it must be when $l_2\geq 0$. If this increase is optimal to make when it would reduce labor supply ($\partial l/\partial\rho<0$), it must be that t_w is negative.

By the same reasoning, if $\partial l/\partial\rho>0$, pushing ρ into a backward-bending stretch of σ would seem to be plausible only if $t_w>0$, so that q_3l rises by less than $F(\cdot)$; but here we have to watch for the possibility that $t_n<0$ so that x_{22} is another "plus" in the case for high ρ. However, we can prove

PROPOSITION B8: *If $\sigma_2\geq 0$ at the optimum, then $l_2t_w<0$ so that t_w and $\partial l/\partial\rho$ are like-signed.*

From (38) it follows that $t_wl_2=-\sigma_2+(1+n)^{-1}x_{22}<0$.

The foregoing analysis may have left the false impression that l_2 can be of any sign *independently* of ρ and ω. In our model, the Hicks-Lucas effect $l_2<0$ requires that first-period leisure and second-period consumption are net substitutes, in the aggregate at any rate, and this substitution effect of a rise of ρ, which encourages less leisure, overcomes the income effect, which encourages more leisure. Our assumptions so far pose no bar to considering the possibility that at least for all $\rho<n$ the income effect is swamped by the substitution effect so that in that region $l_2<0$ (Hicks-Lucas). This leads to the result:

PROPOSITION B9: *If $l_2<0$ for all $\rho<n$, then optimal $l_2<0$. It is then possible that optimal $\rho>n$ and $t_w>0$.*

The proof is by contradiction. If optimal $l_2\geq 0$, optimal $\rho<n$ (Proposition B6). But, by supposition, $l_2<0$ at every $\rho<n$. Hence optimal $l_2\geq 0$ is impossible. As before, optimal $l_2<0$ leaves the signs of t_n and t_w indeterminable from (36a)–(36c). However, it is clear from (37) that $t_n>0$ for some $l_2<0$.

Suppose the optimum is indeed such that $t_n<0$ ("machines are subsidized") and $t_w>0$ ("toil is taxed"). Where is the justice in that? The only possible justification for raising ρ beyond the biologic rate n is its Hicks-Lucas stimulus to the supply of output, $F(k, l)$. Undoubtedly many critics of the "socialist" goal of propertylessness, and of the "humanist" opposition to interest income, have had this "model" implicitly in mind: Prohibitions against property holding and heavy taxation of interest *may* so dampen income-earning incentives as to be inefficient in the maximization of any Bergson-type social welfare function.

Finally, we state the following results although the proof will have to be omitted here:

PROPOSITION B10: *If first-period leisure and second-period consumption are "independent" goods or else net complements for all households at the optimum, then* $t_n > 0$ *and* $t_w > 0$.

C. *Optimal Taxes with Optimized Capital*

In the previous section, capital was fixed and the per capita debt is free to be set at its optimal stationary level. The latter is calculable by regarding d as endogenously determined by $(1+n)^{-1}\sigma(\cdot)-k$. In Act 2, as it were, the shoe is on the other foot. The debt per worker is fixed and capital per worker is allowed to realize its optimal level, determined by

$$(39) \quad k = (1+n)^{-1}\sigma(\cdot) - d \equiv \sigma^*(\cdot) - d$$

when evaluated at optimal q_2, q_3, β_1. It may be noticed that the notion of a welfare-maximizing steady-state k is just the Golden Rule exercise with the new wrinkle of a government deficit constraint. But it is with the optimal taxes in this steady state that we are concerned.

In the special case where $V_2^0 = \beta_2 - x_2^0 = 0$, our new problem may be formulated analogously to (32):

$$(40) \quad -nd = \max_{\{q_2, q_3\}} P(q_2, q_3, b(d); d)$$

in which $P(\cdot)$ is defined as the government's algebraic budgetary surplus per worker:

$$(41) \quad P(\cdot) \equiv F[(\sigma^* - d), l] - q_2^{-1}(1-q_2)\sigma^* - q_3 l - \beta_2(1+n)^{-1} - \gamma - \beta_1$$

$P(\cdot)$ is to be evaluated at the maximum $\beta_1 = b(d)$. The maximization in (40) is constrained by $\sigma^* - d = k \geq 0$ and $\sigma^* \geq 0$. In the following analysis it is assumed that an interior solution exists that makes both k and σ^* positive.

The first-order conditions for an interior maximum are

$$(42) \quad P_2 = 0, \qquad P_3 = 0, \qquad P_\beta \leq 0$$

Equations (41) and (42) yield

$$(43a) \quad l_2 t_w + x_{22}(1+r)(1+n)^{-1}t_r = -(1+r)(1+n)^{-1}(x_2 - \beta_2) < 0$$

$$(43b) \quad l_3 t_w + x_{23}(1+r)(1+n)^{-1}t_r = l > 0$$

$$(43c) \quad l_\beta t_w + x_{2\beta}(1+r)(1+n)^{-1}t_r \leq 1$$

where, recalling (30c), $t_r \equiv (r-\rho)(1+\rho)^{-1} \cdot (1+r)^{-1}$ and $r = F_k$.

A less cumbersome way to write (43) uses $\rho = q_2^{-1}(1-q_2)$ and $\omega = q_3$:

$$(44a) \quad \frac{\partial l}{\partial \rho} \cdot (F_l - \omega) + \frac{\partial \sigma^*}{\partial \rho}(F_k - \rho) = \sigma^*$$

$$(44b) \quad \frac{\partial l}{\partial \omega} \cdot (F_l - \omega) + \frac{\partial \sigma^*}{\partial \omega}(F_k - \rho) = l$$

$$(44c) \quad l_\beta \cdot (F_l - \omega) + \sigma^*_\beta (F_k - \rho) \leq 1$$

We state the following results with remarks on the proofs.

PROPOSITION C1: *Neither* $\rho = -1$ *nor* $\omega = 0$ *is optimal.*

This follows from our assumptions that $\sigma^* > 0$ and $\sigma^* - d > 0$.

PROPOSITION C2: *If optimal* $l_3 < 0$, *then* $t_r > 0$.

In (43b), $t_w l_3 - l < 0$ if $l_3 < 0$ because $l + l_3 q_3 > 0$.

PROPOSITION C3: *Either* $t_r > 0$ *or* $t_w > 0$ *or both.*

If $l_3<0$, $t_r>0$. If $l_3\geq 0$, $t_w>0$ or $t_r<0$ by (43b). If $l_3<0$, nothing new follows if also $l_2\geq 0$. But

PROPOSITION C4: *If optimal* $l_2>0$, *then* $t_r>0$.

If $t_w\leq 0$, $t_r>0$ by C3. If $t_w>0$, and $l_2>0$, then $t_r>0$ by (43a).

Section B was greatly simplified by the result that $l_3>0$. It hinged on the result that $\partial[F(k,\ l)-\omega l]/\partial q_3$ is positive. In a symmetrical world, it would be true here that $\sigma_2<0$, that is, $\partial\sigma/\partial\rho>0$. The hitch is apparently that we do not *generally* have $\partial[F(k,\ l)-\omega l]/\partial q_2$ positive. Hence very low q_2 might be repaid from the labor end. Nevertheless, suppose optimal $t_w l_2\geq 0$. Then (43a) yields

$$(45)\quad [q_2-(1+r)^{-1}]x_{22}+x_2-\beta_2 = -\,l_2t_w(1+r)^{-1}(1+n)<0$$

Hence, from (17),

$$(46)\quad \sigma_2=-\,l_2t_w(1+r)^{-1}(1+n) + (1+r)^{-1}x_{22}<0$$

This leads to the result:

PROPOSITION C5: *If optimal* $l_2t_w\geq 0$, $\sigma_2<0$ (*non-backward-bending solution*). *If* $\sigma_2\geq 0$, *then* t_w *and* $\partial l/\partial\rho(\neq 0)$ *are like-signed.*

For completeness, we solve for $r-\rho$ and t_w from (44a) and (44b):

$$(47)\quad r-\rho=\frac{l\dfrac{\partial l}{\partial\rho}-\sigma^*\dfrac{\partial l}{\partial\omega}}{\dfrac{\partial l}{\partial\rho}\dfrac{\partial\sigma^*}{\partial\omega}-\dfrac{\partial l}{\partial\omega}\dfrac{\partial\sigma^*}{\partial\rho}}$$

$$t_w=\frac{\sigma^*\dfrac{\sigma\sigma^*}{\partial\omega}-l\dfrac{\partial\sigma^*}{\partial\rho}}{\dfrac{\partial l}{\partial\rho}\dfrac{\partial\sigma^*}{\partial\omega}-\dfrac{\partial l}{\partial\omega}\dfrac{\partial\sigma^*}{\partial\rho}}$$

A nominally identical "formula" for t_w can be obtained for the problems posed in Section A and Section B. Analysis of (47) adds nothing to what has preceded.

D. *Optimal Taxes with Optimized Capital and Debt*

When the debt is free, the problem of optimal taxation is to maximize V^0 subject to $J(\cdot)=nk$. Because $V^0(q_2,\ q_3,\ \beta_1,\ \beta_2)$ is not a function (directly) of k, the latter influences maximum $V^0(\cdot)$ only through its impact on the investment constraint. If $F_k-n>0$, an increase of k increases maximum feasible β_1 for any q_2, q_3, β_2. Hence at optimal k, $F_k=n$. If $F_k\neq n$, some change of k could permit higher β_1, hence greater $V^0(\cdot)$ for fixed q_2, q_3, and β_2, so that k could not be the best sustainable, steady k.

The argument is easily formalized for the special case using (27) of Section A. Optimized debt implies

$$(48)\qquad 0=b_d(k,\ d)=\frac{\lambda(1+n)}{-J_\beta+\lambda\sigma_\beta}$$

Hence $\lambda=0$. Also optimal k implies

$$(49)\quad 0=b_k(k,\ d)=\frac{F_k-n+\lambda(1+n)}{-J_\beta+\lambda\sigma_\beta}$$

Therefore when both k and d are optimized

$$(50)\qquad F_k(k,\ l)=n$$

The Golden Rule path, unconstrained by a nonoptimal debt requirement, has a technological characterization in terms of the *capital-labor ratio* k/l, similar to the Golden Rule path's characterization in *technocratic* models where no attention is paid to the institutions, fiscal or otherwise, needed to sustain that path and other parallel, steady-state paths.

Is this result surprising? In one respect, no. To maximize V^0 and β_1, we desire efficiency in production; the Golden Rule path is, in a sense, the *most* "efficient" of all sustainable paths. It leaves the largest slack for consumption per head, and thus for β_1 which "absorbs" this slack.

But in another respect it is surprising and a little misleading. One might have thought that a larger capital stock would impose larger "deadweight" frictional costs beyond the increase in steady-state investment nk; perhaps higher tax *rates* t_w and t_n eventually might be needed to ensure the increase in nk as k approached its technocratic level because effective labor per head l is smaller than what one supposes it would be if taxation were (optimally) lump sum. Actually, institutions do indeed affect the magnitudes of k and l on the Golden Rule path. But they do so equiproportionately. Thus they leave $F_k=n$ and $F_l=\Psi(F_k)=\Psi(n)$, the function Ψ describing the factor price frontier.

With regard to the Golden Rule debt level, if optimal σ^* is less than (greater than) Golden Rule k, then optimal d is negative (positive). To analyze the optimal tax mix, we use $n=r$ and (36a)−(36c) to obtain

$$l_2 t_w + x_{22} t_r = -(x_2 - \beta_2) < 0 \tag{51a}$$

$$l_3 t_w + x_{23} t_r = l > 0 \tag{51b}$$

$$l_\beta t_w + x_{2\beta} t_r = 1, \qquad t_w = \Psi(n) - q_3 \tag{51c}$$

Obviously, all the propositions about $n-\rho$ in Section B hold here, and hold also for $r-\rho$. The propositions in Section C where applicable also hold. We consolidate these as follows.

PROPOSITION D1: *Neither* $\rho=-1$ *nor* $\omega=0$ *is optimum.*

PROPOSITION D2: *At the optimum,* $l_3>0$ *and max* $(t_r, t_w)>0$.

PROPOSITION D3: *If* $\sigma_2>0$ *(non-forward-rising),* $l_2\neq 0$. *If* $l_2<0$, $t_w>0$. *If* $l_2>0$, $t_w<0$ *and* $t_r>0$. COROLLARY: *If* $l_2 t_w \geq 0$, $\sigma_2<0$.

PROPOSITION D4: *If* $\sigma_2<0$, *then* $t_r>0$ *if* $l_2\geq 0$.

PROPOSITION D5: *If* $l_2<0$ *for all* $\rho<n$, *which implies strong net substitutability, then optimal* $l_2<0$ *and* t_r *may be of either sign. If* $l_2\geq 0$ *for all* $\rho>n$, *because of a strong income effect or weak enough substitutability, then optimal* $t_r>0$ *and* l_2 *of either sign.*

PROPOSITION D6: *If, for every household,* x_0^m *and* x_2^m *are either net complements or independent goods, then* $t_w>0$ *and* $t_r>0$.

III. Concluding Remarks

Less than the usual purpose would be served by a recapitulation of the results obtained in this paper. If our paper is important, it is because of the suggestiveness of the model constructed and the learning experience of analyzing it, not the results finally wrung from it.

The model studied here has some obvious limitations that need no comment: the closedness of the economy to international trade, investment, and migration, the perfectness of the capital market, the fixity of the technology, and the constancy of the population growth rate.

Another limitation is the imposition onto the model of steady-state behavior. Clearly our steady states are only long run. In the immediate present, at the outset of the just era, it cannot generally be supposed that the old wealth owners faced after-tax factor rewards equal to the rewards it will be optimal to offer the young generation—whether or not it is desired to maintain capital and public debt per worker. The kind of steady state studied here is thus to be understood as a rest point to which the system will gravitate once intergeneration justice is in force—at least under basically stationary conditions on the technology and population.

Yet our steady-state formulations are not quite suited to this notion of a just rest point. As a lesson in optimal taxation, the problem posed in Section A is underrestricted—for not *every* steady (k, d) state could be a just rest point. If $b_k(k, d) < 0$ or $b_d(k, d) > 0$, for example, one would work off the excess capital or make up the deficiency in debt. On the other hand, Sections B and C are misrestricted—for neither $b_d(k, d) = 0$ nor $b_k(k, d) = 0$ is an optimal rest point except in the singular Golden Rule case where both equalities hold. And it is perhaps obvious that Section D (the Golden Rule state) is too restrictive in one sense—for intergeneration justice should not be assumed to demand that society from every imaginable initial state should trudge its way to the Golden Rule state.

These propositions (or their analogues) are demonstrated in an as-yet preliminary paper by Phelps and John Riley that studies the dynamics of capital and wealth in an economy (similar to the one here save for homogeneity of the population) programmed for Rawlsian *inter*generational justice. However, in neglecting population heterogeneity that paper omits the matter of *intra*generational justice that is the focus here. Perhaps the union of these two papers will sometime bear fruit in a third paper, that one on taxation for Rawlsian justice within and across generations.

REFERENCES

P. A. Diamond, "National Debt in a Neoclassical Growth Model," *Amer. Econ. Rev.*, Dec. 1965, *55*, 1125–50.

K. Hamada, "Lifetime Equity and Dynamic Efficiency on the Balanced Growth Path," *J. Publ. Econ.*, Nov. 1972, *1*, 379–96.

J. R. Hicks, *Value and Capital*, New York 1946.

O. Lange and F. H. Taylor, *On the Economic Theory of Socialism*, Bloomingdale 1938.

R. E. Lucas and L. A. Rapping, "Real Wages, Employment and Inflation," *J. Polit. Econ.*, Sept./Oct. 1969, *77*, 721–54.

K. Marx, "A Critique of the Gotha Program," *Neue Zeit* 1895; reprinted in R. C. Tucker, ed., *The Marx/Engels Reader*, New York 1972.

J. A. Mirrlees, "An Exploration in the Theory of Optimal Income Taxation," *Rev. Econ. Stud.*, Apr. 1971, *38*, 175–208.

J. A. Ordover, "Distributive Justice and Optimal Taxation of Wages and Interest in a Growing Economy," *J. Publ. Econ.*, forthcoming.

——— and E. S. Phelps, "Proportionate Steady-State Taxation of Wealth and Wages for Lifetime Economic Justice," disc. pap. no. 7302, dept. of econ., Columbia Univ., Nov. 1973.

E. S. Phelps and J. G. Riley, "Rawlsian Growth: Dynamic Programming of Capital and Wealth for Intergenerational 'Maximin' Justice," working pap., dept. of econ., Univ. California, Los Angeles, Aug. 1974.

J. Rawls, *A Theory of Justice*, Cambridge, Mass. 1971.

E. Sheshinski, "The Theory of Optimal Income Taxation," Harvard disc. pap., Harvard Univ., Feb. 1971.

Portfolio Equilibrium and the Balance of Payments: A Monetary Approach

By JACOB A. FRENKEL AND CARLOS A. RODRIGUEZ*

The purpose of this paper is to analyze the adjustment mechanism in a small fully employed economy which trades in commodities, securities, and money. We attempt to shift the analysis from exchange considerations that dominate barter theory to those features that are relevant to intertemporal trade, the accumulation of real and financial assets, and their reflection in the balance of payments. The basic characteristics of the model are the specifications of the production technology, the consumption function, and the portfolio of assets.

The structure of the model and the equilibrium of the closed economy are described in Section I. The effects of opening the economy to international trade are analyzed in Section II, and the dynamics of the open economy, the accumulation of assets and capital are analyzed in Section III. In Section IV we analyze the process by which the economy adjusts to monetary changes. In these sections we distinguish between short-run and long-run effects and explore the exact channels—in terms of the various balance-of-payments accounts—through which the adjustments take place. An analysis of changes in the exchange rate is provided in Section V which concludes the article with some empirically testable hypotheses.

I. The Model of the Closed Economy

Consumption is assumed to depend on the real value of assets a.

$$(1) \qquad c = c(a); \qquad c'(a) > 0$$

The real value of assets in terms of consumption is defined as

$$(2) \qquad a \equiv p_k k + m$$

where k is the stock of capital, $m(=M/P)$ the real value of cash balances, M the nominal stock of money, P the money price of consumption, and p_k the relative price of capital in terms of consumption.

Population (the labor force) is assumed constant and all quantities are measured in per capita terms. Although the model we use is a one-sector model in which output and capital are homogeneous, we follow the standard assumption that capital cannot be dismantled. This, together with the assumption that there are "installation" costs, implies that the price of a machine, or equivalently the price of a title to the ownership of a machine, need not equal the price of a unit of output. The number of units of capital is identically equal to the number of titles of ownership. Households are assumed to own the titles and to rent the machines to firms. Their earnings are the returns on the machines.

Output is assumed to be produced by a linear homogeneous production function so that output per capita can be written as $f(k)$ with the usual restrictions.

The investment function embodies a variant of the Penrose effect (see Hirofumi Uzawa and John Gould). Since the industry is constrained by its managerial ability which is assumed to be constant,

---

* Associate professors of economics, University of Chicago and Columbia University, respectively. In revising the paper we have benefited from useful suggestions from George Borts, Rudiger Dornbusch, Michael Schmid, and an anonymous referee.

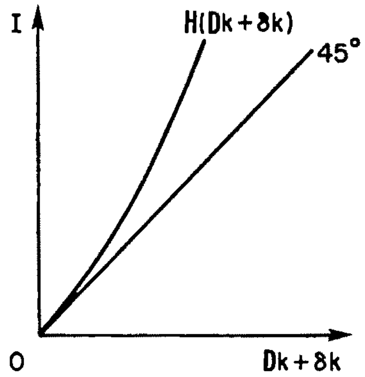


FIGURE 1A

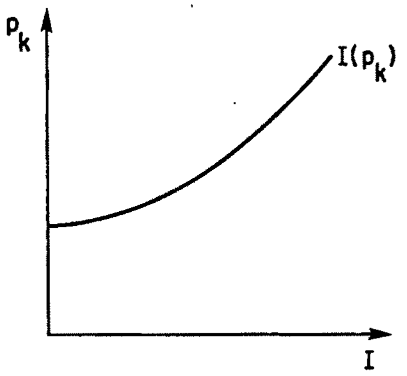


FIGURE 1B

the process of transforming output into nstalled real capital is costly. The adjustnent cost feature of the technology calls or a distinction between investment and gross capital formation. Specifically, deine investment I as the number of units of output that is used in the process of gross capital formation G where $G \equiv Dk + \delta k$, D denotes the operator d/dt, i.e., $Dk \equiv dk/dt$, and δ is the (constant) percentage rate of depreciation. Thus, I can be viewed as the input requirement for gross capital formation. It is assumed that the latter is subject to increasing marginal cost and therefore, that the rate of investnent is related to the rate of gross capital ormation according to

$$I = H(G) \tag{3}$$

where

$$H'(G)>0,\quad H''(G)>0,\quad H(0)=0,\quad H'(0)=1$$

Maximization of profits implies that $p_k = H'(G)$ and therefore the optimal rate of gross capital formation depends on the relative price of capital according to:[1]

$$G \equiv Dk + \delta k = h(p_k) \tag{4}$$

where $h'(p_k) = 1/H''(G) > 0,\ h(1)=0$

Using (4) in (3), results in the rate of investment as a function of p_k:

$$I = H(h(p_k)) = I(p_k) \tag{5}$$

where $I'(p_k) = H'(G)/H''(G) > 0,\ I(1)=0$

Figures 1A and 1B describe equations (3) and (5), respectively. The fundamental aspect that is revealed in these figures is the rising marginal cost (in terms of units of output that are used in the process) of gross capital formation. Note that the same information could be conveyed by replacing Figure 1A with its inverse so as to describe $Dk+\delta k$ as a function of I with diminishing marginal productivity.[2]

---

[1] To derive this relationship, define total output by $Q=F(K,L)$. Output can be used either for consumption or for investment. The firm's revenue is composed of sales of output net of investment and of sales of new equities resulted from net capital formation (p_kDk). Maximization of profits: $F(K,L)-I+p_kDk-wL-rK$ is done subject to $I=H(DK+\delta K)$ where w and r are, respectively, the wage rate and the rental rate (net of capital maintenance) in terms of output. The first-order conditions imply that $p_K=H'(G)$ which results in equation (4) above (see also the Appendix). In order to have a meaningful aggregate of the capital stock, it is assumed that managers (that are in fixed supply for the industry) are mobile among firms so that their rents are equalized across firms.

[2] The production side of the economy could alternatively be described as a two-sector model where the consumption good is produced by a linear homogeneous function of capital and labor and the installed investment good is produced by a linear homogeneous function of the consumption good and of managerial ability. The fundamental element that is required for a determinate demand for investment is diminishing returns in the economy. In the closed economy it is inconsequential whether these diminishing returns are internal to the

The first equilibrium condition is that total output is demanded:

$$(6) \qquad c(a) + I(p_k) = f(k)$$

The second equilibrium condition involves the composition of the portfolio of assets. The desired ratio l of real cash balances to the real value of capital (equities) is assumed to depend on the rate of interest ρ:

$$(7) \qquad m = l(\rho)p_k k \qquad l'(\rho) < 0$$

where the rate of interest is identified with the net rate of return on capital:[3]

$$(8) \qquad \rho = \frac{f'(k)}{p_k} - \delta$$

In equilibrium output is demanded (flow equilibrium (6)), and the actual stocks of assets are held in the desired proportion (stock equilibrium (7)). Short-run equilibrium requires that conditions (6)–(7) are met at each instant. Since at any instant the stocks of capital k and nominal balances M are given, equilibrium in the short run determines the money price of output P and the relative price of capital p_k (and thereby the rate of interest ρ). The effect of changes in the stock of capital on the price of capital (and thus on the rate of interest) can be ascertained by solving for p_k as a function of k from the short-run equilibrium conditions (6)–(7). From (4) the evolution of the capital stock is described by

$$(9) \qquad Dk = h(p_k) - \delta k$$

and in the steady state net capital formation is zero ($Dk=0$) and gross investment is just sufficient to replace the depreciation of capital.

Assume that the closed economy converged to its steady state and denote by $\bar{k}$ and $\bar{p}_k$ the steady-state values of the closed economy stock of capital and price of capital.[4] Associated with the steady-state values of $\bar{k}$ and $\bar{p}_k$ there is (from (8)) the steady-state rate of interest $\bar{\rho}$ and (from (7)) the steady-state value of real balances $\bar{m}$. Given the nominal money stock $\overline{M}$, the steady-state price level is $\overline{P}$.

II. Opening to International Trade

Suppose now that the economy is at the steady state, and is opened to international trade in both securities and goods. Suppose that the economy is small relative to the rest of the world so that it faces a given world money price of goods P^* and a given world rate of interest ρ^*. Without loss of generality it is assumed that $\rho^* < \bar{\rho}$.

To accommodate the analysis of the open economy under a pegged exchange rate, the closed economy model in the preceding section must be modified in some respects. First the authorities must determine the relative price of the foreign currency in terms of the domestic currency (the exchange rate). It is assumed that upon the opening to trade, the exchange rate e is set according to the purchasing power parity so that

$$(10) \qquad e = \overline{P}/P^*$$

---

firm or arise from the idiosyncracy of production processes; in the open economy a discrepancy is possible between the production and installation of capital goods, and thus a determinate demand for each activity cannot be derived from the standard two-sector model. An alternative derivation of the demand for investment is provided by Stanley Fischer and Frenkel (1972).

[3] We assume that expectations are static and thus we do not allow for anticipated capital gains or losses on equity holdings. Similarly, we assume that the expected rate of inflation is zero and thus we do not distinguish between the real and the nominal rates of interest. Since when the small economy is open it is assumed to face a fixed world price level, we prefer, for the sake of simplicity, not to complicate the structure of the closed economy model by incorporating inflationary expectations. We have analyzed the dynamics of inflation in a similar closed economy model in our 1975a paper.

[4] It can be shown that both the capital stock and its relative price approach the steady state monotonically. The system must be stable if a rise in the stock of capital lowers its price. If, however, a rise in the capital stock induces a rise in its price, the system may still be stable if the rate of depreciation δ is high.

and thus the price of goods in terms of domestic currency remains unchanged when trade opens. Second, we have to distinguish between the location of the stock of capital which determines production and the ownership of capital which determines consumption and portfolio choice. The home country's ownership of capital is denoted by k_D and we continue to denote the stock of capital that is located in the home country by k. If k differs from k_D, foreign ownership (or debt) makes up for the difference. For example, if k were to exceed k_D, foreigners own k_F units of domestically located capital where

$$k_F \equiv k - k_D \tag{11}$$

When the home country is a net debtor ($k_F > 0$), it is assumed that it does not own any foreign equities and when it is a net creditor ($k_F < 0$), it owns the entire domestic capital stock plus $k_D - k$ units of foreign equities. Accordingly, the definition of domestically owned assets becomes

$$a \equiv p_k k_D + m \tag{2'}$$

and the liquidity preference function becomes

$$m = l(\rho) p_k k_D \tag{7'}$$

Short-run equilibrium of the open small economy requires that the money price of goods in terms of each currency is equalized (equation (10)) and that the domestic rate of interest equals the world rate. The latter condition implies that

$$\rho = \frac{f'(k)}{p_k} - \delta = \rho^* \tag{12}$$

Thus, upon the opening to trade (with $k = \bar{k}$), the domestic price of capital rises to p_k' (as a result of the excess demand for domestic securities) so that $[f'(\bar{k})/p_k'] - \delta = \rho^*$. Thus, the opening to trade results in a capital gain on the ownership of domestic capital and consequently in a higher value of assets.[5] The decline in the rate of interest raises the desired ratio of real balances to capital (equities) which in turn induces an instantaneous adjustment of the portfolio. These adjustments are shown in Figure 2 where the pretrade equilibrium is at point A, the initial holdings of capital and real cash balances are $\bar{k}$ and $\bar{m}$, respectively, the real value of assets is $\bar{a} = \bar{m} + \bar{p}_k \bar{k}$, and the pretrade price of capital is $\bar{p}_k$. Upon the opening of trade the rate of interest falls to ρ^*, the price of capital rises to p_k', and the value of assets rises to $a' = \bar{m} + p_k' \bar{k}$. Given the lower rate of interest and the higher value of assets, short-run equilibrium is attained at point B (the intersection of the liquidity preference schedule (7') with the new asset constraint), and the holdings of equities and real cash balances are changed to k_D' and m', respectively. Accordingly, upon the opening to trade the economy exchanges with the rest of the world $\bar{k} - k_D'$ units of equities for $m' - \bar{m}$ units of real cash balances as portfolio equilibrium moves instantaneously from point A to point B.

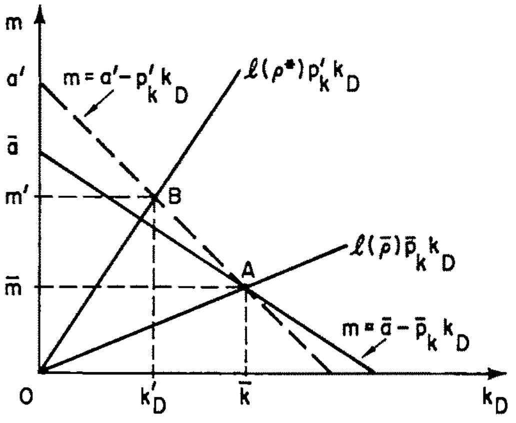


FIGURE 2

---

[5] The effects of these capital gains on the gains from trade in a two-country model are analyzed in Fischer and Frenkel (1974).

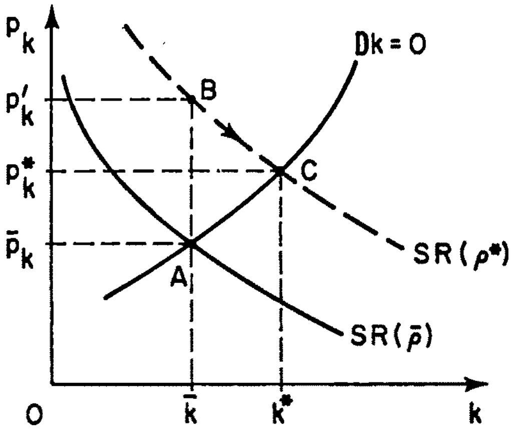


FIGURE 3

III. Dynamics of the Open Economy

The higher price of capital increases the profitability of investment and thus results in a higher equilibrium steady-state capital stock. The dynamics of the capital stock and of the price of capital are described in Figure 3. The locus of short-run equilibria SR describes combinations of p_k and k which maintain a constant rate of interest[6] (a lower rate of interest corresponds to a higher locus in Figure 3). The $(Dk=0)$ locus is derived from equation (9) and describes combinations of p_k and k for which the capital stock remains constant. Its slope is positive since from (4) a higher price of capital increases profitability of gross capital formation while a higher stock of capital increases the depreciation requirement.

Upon the opening to trade, short-run equilibrium moves from point A to point B as the price of capital rises to p'_k. At point B, $Dk>0$ and the system moves monotonically along the $SR(\rho^*)$ schedule towards the new steady state (point C) at which $k=k^*$ and $p_k=p^*_k$. As can be seen, the new steady-state price of capital must exceed $\bar{p}_k$ so as to induce the larger gross investment that is required to just replace the depreciation on the higher stock of capital.

Since consumption depends on the value of assets, we need to analyze the dynamics of assets in order to characterize the evolution of the economy through time. We turn now to the dynamics of the value of assets and then proceed to analyze the dynamics of the balance of payments and its components.

The rate of change in the value of assets is comprised of planned additions to assets (planned savings) and of capital gains and can be written as:

$$Da = [f(k) + p_k Dk - I(p_k) - \rho^* p_k k_F - c(a)] + k_D Dp_k \tag{13}$$

The interpretation of (13) is as follows: The first three terms on the right-hand side of (13) correspond to net *domestic* product (*NDP*) while the first four terms include international dividend payments and correspond to net *national* product (*NNP*). The bracketed term corresponds, therefore, to *ex ante* savings. The *actual* change in the value of assets Da, is an *ex post* concept which corresponds to the sum of *ex ante* savings plus capital gains on equity holdings $(k_D Dp_k)$.[7]

Since all the variables in (13) depend on k and a alone,[8] we can write the dynamic equations for the capital stock and the value of assets as

$$Dk = Dk(k) \tag{9'}$$

$$Da = Da(k, a) \tag{13'}$$

It is shown in the Appendix that around

---

[6] From (8), the equation of the SR locus is $p_k=f'(k)/(\rho+\delta)$ and its slope is $(dp_k/dK)_{SR}=f''/(\rho+\delta)<0$.

[7] This distinction between intended and actual savings presumes that all unanticipated capital gains are absorbed in asset accumulation while all desired consumption plans are fulfilled.

[8] From (2) and (7) $k_D=a/[p_k(1+l)]$; from (11) $k_F=k-k_D$ and from (9) Dk depends on p_k and k. From (12)—given the rate of interest—p_k depends only on k (see fn. 6) and therefore $Dp_k=[f''(k)/(\rho+\delta)]Dk$. Thus all variables in (13) depend on k and a alone.

the steady state, $\partial Da/\partial k > 0$.[9] Since $\partial Dk/\partial k < 0$, stability requires that around the steady state $\partial Da/\partial a < 0$. Intuitively, the rise in the value of assets raises consumption according to the consumption function, and raises income by the yield on the additional assets; the stability condition assures that the induced rise in consumption exceeds the additional yield on assets so as to prevent a further asset accumulation.

The dynamics of the capital stock and the value of assets following the opening to trade are analyzed in Figure 4. Combinations of k and a for which k and a are constant are shown by the $(Dk = 0)$ and the $(Da = 0)$ loci, respectively. The slopes of these loci (around the stable steady state) are determined by (9′) and (13′), and the arrows in Figure 4 correspond to the signs of the partial derivatives. Upon the opening to trade the capital stock is $\bar{k}$ and since $\rho^* < \bar{\rho}$, the new steady-state capital stock must exceed $\bar{k}$ and thus the $(Dk = 0)$ locus must be to the right of $\bar{k}$. The initial value of assets immediately after trade opens depends on the extent of the capital gain that is induced by the rise in the price of capital. In fact the rise in the value of assets may raise consumption by an amount that is sufficient to cause dissavings as shown in the path starting from B in Figure 4. Depending on the extent of the capital gain, the initial point might be above the $(Da = 0)$ locus in which case initially $Da < 0$ as in points B and B', or it might be below the $(Da = 0)$ locus in which case $Da > 0$ as in point B''. Whatever is the actual path, the economy converges eventually towards the new steady state (point C in Figure 4) in which k^* and a^* are the open economy's steady-state values of capital and assets.

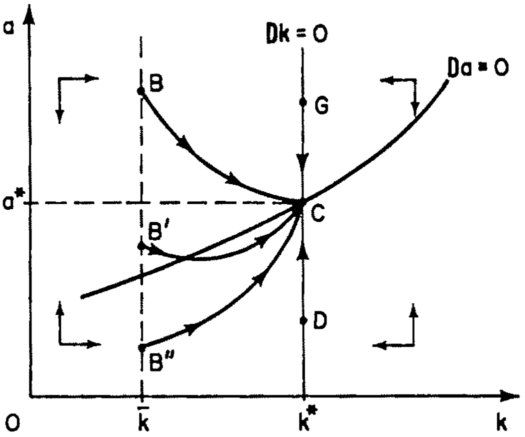


FIGURE 4

The surplus in the balance of trade b_T is defined as the excess of domestic output over absorption (consumption plus investment):

$$(14) \qquad b_T \equiv f(k) - c(a) - I(p_k)$$

The budget constraint implies that the overall surplus in the balance of payments Dm equals the sum of the surpluses in the trade account b_T, the service account $(b_S = -\rho^* p_k k_F)$, and the capital account $(b_C = p_k D k_F)$:[10]

$$(15) \qquad b_T = Dm + \rho^* p_k k_F - p_k D k_F$$

The surplus in the current account corresponds to the excess of income over expenditures, i.e., to the accumulation of assets. Under the assumption that the economy is not growing, the steady-state value of assets is fixed and thus the current account is balanced $(b_T + b_S = 0)$. Since the current account is balanced, the surplus in the capital account b_C must equal the overall surplus in the balance of payments Dm. In the absence of autonomous domestic credit creation, both the capital account and the overall surplus in the balance of payments are zero. Whether the steady-state service account is in surplus or in deficit depends on the balance of indebtedness, i.e., on whether the economy is a net creditor or a net debtor. This cannot be determined on a priori grounds since it depends on two independent considerations—that of savings and that of investment. If the cumulative acquisition of equities k_D exceeds the cumulative supply of domestic equities k^*, then the economy is a net creditor and the steady-state service account is in surplus, and conversely when k_D falls short of k^*.

---

[9] An intuitive explanation for that sign is as follows: a rise in the capital stock (holding assets constant) raises k_F and does not raise disposable income since around the steady state the capital stock is at its optimal level and the rise in output is just sufficient to service the larger debt. Consequently, planned savings do not change and the only effect on the rate of change in the value of assets arises from capital gains. Since around the steady state $\partial Dp_k/\partial k = [f''(k)/(\rho+\delta)]\partial Dk/\partial k > 0$, it follows that $\partial Da/\partial k > 0$.

[10] For a similar framework for the analysis of the balance of payments accounts see Rudiger Dornbusch (1971, 1975), Frenkel (1971, 1975a), Frenkel and Fischer, Ryutaro Komiya, and Douglas Purvis.

IV. The Adjustments to Monetary Changes

Consider a once and for all change in the monetary stock which is brought about by an unanticipated governmental transfer. Given the exchange rate, money prices are given and therefore changes in the nominal stock of cash balances correspond to changes in the real stock. Suppose that at time t_0 the real stock of cash balances is raised by dm_0.[11] The impact effect of this change is to raise the value of assets from a^* to a^*+dm_0 and thus (given the rate of interest) to induce an excess demand for equities that is equal in value to $p_k^* dk_D = dm_0/(1+l)$. Accordingly, instantaneous portfolio adjustment results in an immediate exchange of part of the increased monetary stock for equities resulting therefore in larger holdings of both real cash balances and equities.

Figure 5 describes the impact effect of this change where initial equilibrium is at point A and the new instantaneous equilibrium is attained at point B with equilibrium holdings of real cash balances that are higher by $dm = dm_0 - dm_0/(1+l)$.

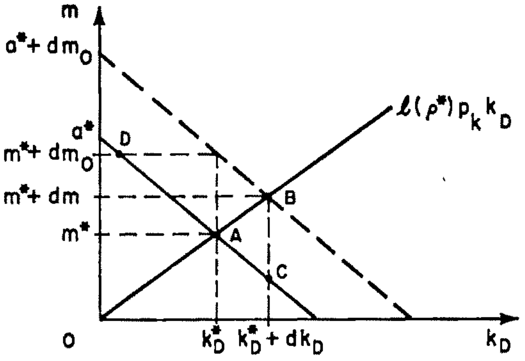


FIGURE 5

In terms of Figure 4 the impact effect is described by moving the economy from point C to point G along the $(Dk=0)$ locus. Therefore, starting at point G, the dynamics of the adjustment is characterized by a continuous reduction in assets until the original steady state is attained at point C in Figure 4. Since the monetary change does not affect the stock of capital and its price, it does not affect the optimal rate of capital formation and thus the path of adjustment must be along the $(Dk=0)$ locus. It follows that along the path the capital stock and the price of capital remain constant at their previous levels k^* and p_k^*.

The instantaneous adjustment of the portfolio as well as the transition to the new steady state are effected through the various accounts of the balance of payments. Since the monetary change does not affect the capital stock, the level of output is fixed and consumption rises. Therefore, the impact effect on the balance of trade is to lower the surplus (or raise the deficit) by $(\partial c/\partial a)dm_0$. The instantaneous adjustment in the portfolio of assets involves an exchange of a fraction $1/(1+l)$ of the increased money stock for securities. This exchange improves the balance of indebtedness and causes an improvement in

---

[11] In fact, the exogenous change in the nominal money stock is dM_0; however, since the price level is given, this change in the nominal stock corresponds to a change in the real stock by dm_0.

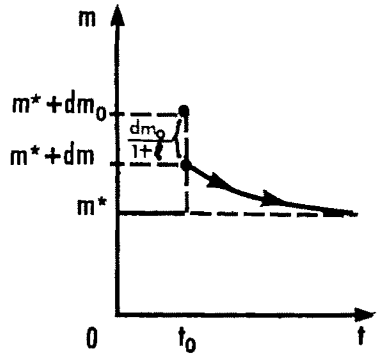


FIGURE 6A

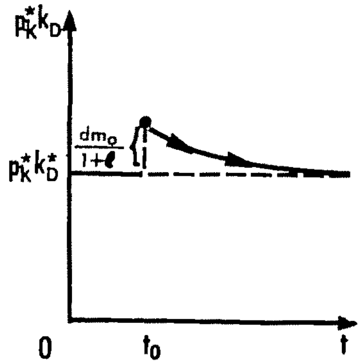


FIGURE 6B

the service account by $\rho^* dm_0/(1+l)$. The sum of the impact effects on the balance of trade and on the service account is the impact effect on the current account, i.e., on asset accumulation. When the system is stable, $\partial Da/\partial a = -\partial c/\partial a + \rho^*/(1+l) < 0$, and therefore the impact effect is to induce instantaneous asset decumulation at the rate $(\partial Da/\partial a)dm_0$. This desired asset decumulation will be divided between securities and money so as to maintain portfolio balance. Consequently, *aside from the initial stock adjustment*,[12] the impact effects will be a surplus of $-(\partial Da/\partial a)dm_0/(1+l)$ in the capital account and a deficit of $l(\partial Da/\partial a)dm_0/(1+l)$ in the overall balance of payments.

So far we have examined the impact effects of the monetary change. Accordingly, the effects on the various accounts of the balance of payments were confined to the new instantaneous short-run equilibrium when the economy is at point G in Figure 4 and at point B in Figure 5. We turn now to the analysis of the paths of the various accounts in the transition to the steady state.

When the value of assets exceeds the steady-state value, individuals run down their holdings of both equities and real cash balances and thus induce a surplus in the capital account and a deficit in the monetary account (the overall balance of payments). The stock of assets changes through time according to $(\partial Da/\partial a)Da < 0$ and this path corresponds to the path of the current account $Db_T + Db_S$. Since during the adjustment period assets decline, consumption falls and the balance of trade improves. On the other hand, since along the path the economy is exporting securities, its indebtedness increases through time by $Da/(1+l)$ and thereby, the service account deteriorates by $\rho^* Da/(1+l)$. The surplus in the capital account and the deficit in the overall balance of payments are proportional to Da and decline through time. The process comes to an end when the initial asset position is restored and $Da = 0$.

The paths of the stocks of money and ownership of equities are described in Figure 6, while Figure 7 summarizes the

---

[12] Since our analysis is conducted in terms of continuous time, the algebraic formulation of the exact paths is complicated due to the fact that there are both adjustments that occur through time and adjustments that occur instantaneously in attaining a continuous portfolio balance. We have defined the capital account and the overall balance of payments in terms of the changes in the holdings of securities and money ($p_k Dk_F$ and Dm) that occur *through time*; therefore, these definitions do not capture international transactions that occur discretely as those that are involved in the instantaneous exchange of $dm_0/(1+l)$ of cash balances for equities. Since empirical data are reported discretely, this transaction will be reported as a deficit in the overall balance of payments and an equal deficit in the recorded capital account. The following expressions that pertain to the impact effect do not include these initial stock adjustments and correspond therefore to the situation at point B in Figure 5.

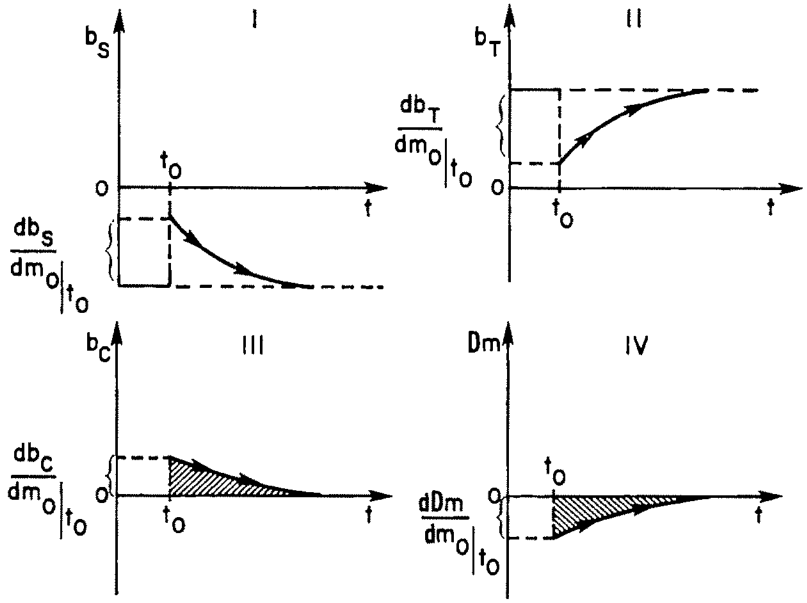


FIGURE 7

dynamics of the various balance of payments accounts following the once and for all rise in the monetary stock.

In terms of Figure 5, the transition toward the steady state is characterized by a movement from B to A along the portfolio balance locus. Since during the transition period portfolio equilibrium obtains, it follows that the percentage rate of change in the monetary stock must equal the percentage rate of change in the holdings of equities. Therefore at each point of time along the path (following the initial restoration of portfolio balance), the ratio of the deficit in the overall balance of payments to the surplus in the capital account must be l.

The adjustment period continues as long as the value of assets exceeds a^*. The cumulative deficit in the overall balance of payments is the shaded area in Panel IV of Figure 7 and is equal to the initial rise in the equilibrium holdings of money $[dm_0 - dm_0/(1+l)]$. Likewise, the cumulative surplus in the capital account is the shaded area in Panel III of Figure 7 and is equal to the initial rise in the value of the equilibrium holdings of equities $[dm_0/(1+l)]$. The sum of the two shaded areas is the initial rise in the quantity of money dm_0 and the ratio of the two areas is equal to the desired ratio of money to equities l.

To summarize, changes in the quantity of money which are brought about by a transfer do not affect the long-run equilibrium of the system—a result which is common to the literature on the monetary approach to the balance of payments.[13] The adjustment to such monetary changes takes two forms. First, there is an instantaneous adjustment in the portfolio of assets in which part of the increased stock of money is immediately exported in exchange for equities. Second, there is a period of adjustment during which the economy runs down assets.[14] These adjustments are reflected in the various accounts of the balance of payments and the balance of indebtedness.

---

[13] For a collection of contributions to this literature, see Frenkel and Harry Johnson. See also Frenkel (1975b).

[14] A similar distinction between the instantaneous part and the time-dependent part of the adjustment process is emphasized by Dornbusch (1975a), Michael Mussa, and Purvis.

Consider next the effects of changes in the quantity of money which are brought about through open market operations in equities. When the government owns part of the capital stock k_G, equation (11) becomes

$$k = k_D + k_G + k_F \tag{11'}$$

When the quantity of money is raised by dm_0 through an open market purchase then

$$dm_0 = \overset{*}{p_k} dk_G = - \overset{*}{p_k} dk_D \tag{16}$$

The effects of these operations depend on the manner by which the government disposes of the returns from their holdings of securities as well as on the manner by which the private sector treats these governmental actions. It is assumed that the returns on governmental holdings of equities are returned to the public as transfers or tax rebates.

In analyzing the reaction of the private sector to the open market purchase, it is useful to consider two extreme cases. Assume first that the private sector does not capitalize the flow of transfers. In this case an open market purchase does not affect the wealth position of individual asset holders, but it raises the ratio of real balances to equities above the desired ratio l. In terms of Figure 5, it moves the current holdings of assets from point A to point D. Since the rate of interest is given, individuals will restore portfolio equilibrium instantaneously (to point A) by exchanging the increased stock of cash balances for foreign securities and thus reducing the economy's net debtor position by $p_k^* dk_F = dm_0$. In other words, the fact that the open market operation is conducted in assets that are traded internationally at a given price enables the private sector to nullify the actions of the monetary authority. In fact, in this case open market operations amount to an exchange of foreign exchange reserves for securities between the monetary authorities and foreign asset holders.[15] Since the economy's net debtor position declines by $p_k^* dk_F$, the instantaneous deficit in the service account must also decline by $\rho^* p_k^* dk_F$. However, since $k_F = k - k_G - a / [p_k^*(1+l)]$, it follows from (13) that an open market purchase shifts the $(Da=0)$ locus upwards to a point like G in Figure 4. Along the path to the new long-run equilibrium the economy imports assets (securities and money), the service account improves, and the balance of trade deteriorates. Long-run equilibrium is reached when the increased value of assets induces higher consumption at a rate that is equal to $\rho^* p_k^* dk_G$, the noncapitalized governmental transfer.

Consider next the other extreme case where the private sector anticipates and capitalizes the flow of transfers $\rho^* p_k^* dk_G$, and assumes that these claims are marketable like any other equity. Effectively therefore, the private sector conceives of the equities that are held by the government as its own equities that are deposited in the governmental safekeeping box. Accordingly, the wealth of the private sector is $a = p_k^*(k_D + k_G) + m$ and the liquidity preference function is $m = l(\rho^*) p_k^*(k_D + k_G)$. An open-market purchase in this case does not affect the sum $(k_D + k_G)$, but it raises the monetary stock by dm_0. Consequently, the effects of an open-market purchase are identical to the effects of an increase in the monetary stock that is brought about through an unanticipated governmental transfer of cash balances.

The analysis of the two extreme cases

---

[15] These implications are elaborated by Robert Mundell (1968, 1971), Ronald McKinnon, Johnson (1972), and Dornbusch (1974, 1975a).

reveals that the differences between the effects of open-market operations and the effects of unanticipated transfers of cash balances stem from issues that are related to the degrees of capitalization and marketability of claims to future income streams.[16]

The effects of changes in the liquidity preference can be easily analyzed by including a shift parameter in the liquidity preference function (7′) and by exploring the exact channels—in terms of the various balance-of-payments accounts—through which adjustment takes place. It can be shown that a rise in the desired ratio of money to equities has no effect on the stock of capital and on its price, while it results in lower long-run holdings of assets. Following the initial stock adjustment (which involves an instantaneous sale of equities for money), the economy gradually runs down assets. The cumulative deficit in the balance of payments may exceed or fall short of the initial instantaneous rise in the monetary stock. In other words, the long-run effects on the holdings of real balances are ambiguous.

V. The Effects of Changes in the Exchange Rate

To analyze changes in the exchange rate recall that the real quantity of cash balances is $m = M/eP^*$. Since the money price of goods in terms of foreign currency is given, it follows that the analysis of changes in the exchange rate is identical to the analysis of changes in the nominal quantity of money. In fact, a devaluation of the currency is exactly an imposition of a capital levy on real cash balances, and therefore the analysis in the previous section (in which the quantity of money was raised through an unanticipated transfer) is identical to the analysis of a revaluation of the currency, while the analogue to a devaluation of the currency is a reduction of the nominal quantity of money. This correspondence is implied by the fact that the system is homogeneous of degree zero in the nominal money stock M and in the exchange rate e. In line with the literature on the monetary approach to the balance of payments, a devaluation must improve the balance of payments whenever it induces an excess demand for money.[17] The additional insight provided by the present analysis is in identifying the exact channels—in terms of the various balance-of-payments accounts—through which the adjustment takes place. The distinction between the impact effect and the transition period emphasizes the role of the portfolio of assets in enabling the private sector to build up instantaneously part of the depreciated real monetary stock. The analysis of devaluation can be shown in terms of Figure 5. Consider an initial long-run equilibrium at point B. A devaluation of the currency lowers the real value of cash balances to point C. An instantaneous adjustment of the portfolio enables the private sector to exchange securities for cash balances and to arrive at point A, and thus to build up immediately part of the depreciated monetary stock. The rest of the adjustment is carried out through time as the community reduces expenditures relative to income and thereby runs a balance of trade surplus. The accumulation of assets along the path between point A and point B is reflected in a deficit in the capital account and a surplus in the overall balance of payments.[18] The process of adjustment is completed once the nominal stock of money has risen equiproportionally to the depreciation of the currency.

---

[16] These issues are analyzed in the context of a closed economy by Lloyd Metzler and Mundell (1971, ch. 1).

[17] See Eitan Berglas, Dornbusch (1973), Johnson (1972, 1975), Mundell (1971), Mussa, and Johan Myhrman.

[18] Similarly, in terms of Figure 4 a devaluation lowers the value of assets from point C to point D and the adjustment is described by the path DC. Since the yield on equities is given by the world market, the capital stock remains constant during the transition and the path of adjustment lies along the $(Dk=0)$ locus. Dornbusch (1975a) has analyzed the effects of devaluation in

In summary, the above analysis of devaluation provides some testable hypotheses:

(i) The *initial* effects of a devaluation should induce a surplus in the capital account, an improvement in the trade account, a deterioration in the debt service account, and a surplus in the monetary account. In the main, the monetary account and the capital account reflect the once and for all *stock* adjustment that is required to restore portfolio balance and in the short run most of the adjustment is effected through these assets accounts.

(ii) During the later phases of the adjustment process, the gradual accumulation of assets is reflected in a deficit in the capital account and a surplus in the monetary account, both of which decline through time as the adjustment process comes close to an end. During these phases the trade account deteriorates while the debt service account improves.[19]

(iii) Since along the path of adjustment portfolio equilibrium obtains, we expect that during the later phases the ratio of the surplus in the monetary account to the deficit in the capital account should approximate the desired composition ratio l.[20]

(iv) Although the devaluation per se does not induce changes in the rate of capital formation, the adjustment process involves changes in net ownership of foreign securities. Consequently we expect that following a change in the exchange rate, the path of *GNP* will display larger variations than the path of gross domestic product (*GDP*).

The model also reveals that if the exchange rate were flexible, then a doubling of the nominal money stock would result in an equiproportionate change in the exchange rate. In other words, the fact that the system is homogeneous of degree zero in M and e implies a symmetry between the pegged and the flexible exchange rate systems. Under a pegged exchange rate system a devaluation of the currency results, in the long run, in an equiproportionate rise in the nominal stock of money, while under a flexible exchange rate system a doubling of the nominal stock of money results in an equiproportionate depreciation of the currency. In both cases, there are no long-run changes in the real value of cash balances as well as in all other real quantities—although of course a devaluation of the currency raises the real stock of international reserves when the exchange rate is pegged.

Although the long-run effects are symmetric under the two exchange rate sys-

---

a model which includes internationally traded bonds in addition to equities that are nontraded assets and therefore their yield is determined endogenously. In such a model a devaluation (from an equilibrium position) induces a transitory cycle of the capital stock.

[19] It is of interest to note that in a recent paper, using a different model, Dennis Logue and Thomas Willett applied a portfolio approach to the analysis of exchange rate changes and conclude that ". . . portfolio rebalancing generally suggests therefore that a devaluation will lead to capital inflow . . . as investors adjust their portfolio to take account of the wealth effects of an exchange rate realignment." This conclusion is in full conformity with our analysis of the *initial* phase of the adjustment—the phase in which desired portfolio *composition* is restored. However, the desired *size* of the portfolio is attained in later phases during which the capital account is in a deficit. Moreover, Logue and Willett assume that the ". . . adjusting country pursues a monetary-fiscal policy mix which leaves aggregate demand unaltered." Our analysis shows that when such policies are not followed, individuals reduce expenditures relative to income in an attempt to rebuild their depreciated stock of assets, and thereby induce an improvement in the trade account.

[20] Strictly speaking, these hypotheses are confined to the case in which the monetary authorities maintain a rigid ratio between the stock of money and international reserves and do not pursue independent credit and sterilization policies. The hypotheses can be easily modified to accommodate the case in which the composition ratio of the Central Bank's assets does change. For a framework for the analysis of the effects of changes in the composition ratio, see Johnson (1972).

tems, the adjustment process is not. When the exchange rate is pegged the recuperation of cash balances following a devaluation is gradual and is effected through time as the community runs a surplus in the balance of payments. When the exchange rate is flexible, the restoration of the equilibrium value of real cash balances following a rise in the nominal stock of money is instantaneous and the exchange rate depreciates immediately at the instant of the monetary change. This latter phenomenon emphasizes the fact that the exchange rate, being the relative price of monies, is determined in a similar way as other relative asset prices and thus the equilibrium exchange rate is that relative price of monies at which the existing stocks are willingly held.[21]

APPENDIX

In this Appendix we derive equation (13′) in the text and derive the signs of its partial derivatives. In deriving the optimal behavior of the firm it is convenient to define quasi profits π as the return to labor plus rents to entrepreneurs. The firm's revenue is composed of sales of output that are not invested $[f(k)-H(Dk+\delta k)]$ and of sales of new equities (p_kDk). Its interest cost on capital is $\rho p_k k$. Thus π is:[22]

$$\text{(A1)} \qquad \pi = f(k) - H(Dk + \delta k) + p_k Dk - \rho p_k k$$

The firm, assuming that p_k is fixed, attains the rate of net capital formation such that

$$\text{(A2)} \qquad \frac{\partial \pi}{\partial Dk} = - H'(Dk + \delta k) + p_k = 0$$

thus,

$$\text{(A3)} \qquad Dk = h(p_k) - \delta k$$

The choice of the optimal capital intensity for the representative firm satisfies (A4):[23]

$$\text{(A4)} \qquad \frac{\partial \pi}{\partial k} = f'(k) - \delta H'(Dk + \delta k) - \rho p_k = 0$$

and using (A2), (A4) becomes

$$\text{(A4}'\text{)} \qquad p_k = \frac{f'(k)}{\rho + \delta} = p_k(k); \qquad \partial p_k/\partial k < 0$$

Using (A4′) in (A3) yields

$$\text{(A3}'\text{)} \qquad Dk = Dk(k); \qquad \partial Dk/\partial k < 0$$

From (A4′) Dp_k can be written as

$$\text{(A5)} \qquad Dp_k = \frac{f''(k)}{\rho + \delta} Dk$$

and using (A3′) yields

$$\text{(A5}'\text{)} \quad Dp_k = \frac{f''(k)}{\rho + \delta} Dk(k) = Dp_k(k)$$

When (A5′) is evaluated around the steady state (with $Dk=0$),

$$\text{(A6)} \qquad \left. \frac{\partial Dp_k}{\partial k} \right|_{Dk=0} = \frac{f''(k)}{\rho + \delta} \frac{\partial Dk}{\partial k} > 0$$

The rate of change in the value of assets can be written (using (13) in the text) as

$$\text{(A7)} \quad Da = \bar{\pi}(k) + \rho^* p_k k_D - c(a) + k_D Dp_k$$

where $\bar{\pi}(k)$ is equation (A1) above, incorporating (from (A4′)) the negative dependence of p_k on k which is ignored by the competitive firm. But since $p_k k_D = a/(1+l)$, we have,

---

[21] This view on the role and determination of the rate of exchange is in sharp contrast with the popular notion that the exchange rate is determined in the flow market so as to assure a balanced balance of payments. The role of stocks rather than flows in exchange rate theory was recognized in some early writings but ignored to a large extent in more recent approaches. The revival of the asset view is a natural application of the monetary approach to the balance of payments; see for example Dornbusch (1975b).

[22] Alternatively (A1) could be written in the form

$$\text{(A1}'\text{)} \qquad \pi = f(k) - H(Dk+\delta k) + p_k(Dk+\delta k) - p_k(\rho+\delta)k$$

where the last term corresponds to the concept of "user cost." Obviously (A1′) is identical to (A1), and therefore reflects the fact the real cost of using capital (and thus optimal use) is independent of contractual arrangements regarding capital maintenance.

[23] Note that to the extent that we assume that capital is specific at the level of the firm, (A4) might be viewed as the condition for optimal allocation of labor among firms. We are indebted to Robert Lucas, Jr. and Hiroyuki Hino for useful discussion on these issues.

using (A5′),

$$\text{(A7}'\text{)}\quad Da = \tilde{\pi}(k) + \frac{\rho^*}{1+l}a - c(a) + \frac{a}{1+l}\,\frac{Dp_k(k)}{p_k(k)}$$

that is,

$$\text{(A8)}\qquad Da = Da(k, a)$$

which is equation (13′) in the text.

We wish to evaluate $\partial Da/\partial k$ and $\partial Da/\partial a$ around the steady state (where $Dk = Dp_k = 0$). From (A7′),

$$\text{(A9)}\quad \frac{\partial Da}{\partial k} = \frac{\partial \tilde{\pi}}{\partial k} + \frac{a}{p_k(1+l)}\,\frac{\partial Dp_k}{\partial k}$$

But since around the steady state the competitive firm equates the marginal productivity of capital $f'(k)$ with user cost $(\rho^*+\delta)p_k$, $\partial\tilde{\pi}/\partial k = -\rho^* k\partial p_k/\partial k > 0$, and the only source for changes in the rate of change of the value of assets is capital gains. Using (A4′) and (A6) in (A9) yields:

$$\text{(A9}'\text{)}\quad \frac{\partial Da}{\partial k} = \left[\frac{a}{p_k(1+l)}\,\frac{\partial Dk}{\partial k} - \rho^* k\right] \cdot \frac{f''}{\rho^*+\delta} > 0$$

From (A7′), around the steady state

$$\text{(A10)}\qquad \frac{\partial Da}{\partial a} = \frac{\rho^*}{1+l} - \frac{\partial c}{\partial a}$$

The rise in the value of assets increases interest earnings by $\rho^* p_k \partial k_D/\partial a = \rho^*/(1+l)$. A rise in assets, in turn, raises consumption by $\partial c/\partial a$, and therefore, stability requires that $\partial c/\partial a > \rho^*/(1+l)$ so as to prevent a further accumulation of assets. Thus, stability requires that $\partial Da/\partial a < 0$.

REFERENCES

E. Berglas, "Devaluation, Monetary Theory and Border Tax Adjustments," *Can. J. Econ.*, Feb. 1974, *7*, 1–11.

R. Dornbusch, "Notes on Growth and the Balance of Payments," *Can. J. Econ.*, Aug. 1971, *4*, 389–95.

——, "Devaluation, Money, and Nontraded Goods," *Amer. Econ. Rev.*, Dec. 1973, *63*, 871–80.

——, "Capital Mobility and Portfolio Balance," unpublished paper, Univ. Chicago 1974.

——, (1975a) "A Portfolio Balance Model of the Open Economy," *J. Monet. Econ.*, Jan. 1975, *1*, 3–20.

——, (1975b) "Flexible Exchange Rates, Capital Mobility and Macroeconomic Equilibrium," in E. Classen and P. Salin, eds., *Recent Issues in International Monetary Economics*, Amsterdam 1975.

S. Fischer and J. A. Frenkel, "Investment, the Two-Sector Model and Trade in Debt and Capital Goods," *J. Int. Econ.*, Aug. 1972, *2*, 211–33.

—— and ——, "Interest Rate Equalization and Patterns of Production, Trade and Consumption in a Two-Country Growth Model," *Econ. Record*, Dec. 1974, *50*, 555–80.

J. A. Frenkel, "A Theory of Money Trade and the Balance of Payments in a Model of Accumulation," *J. Int. Econ.*, May 1971, *1*, 159–87.

——, (1975a) "A Dynamic Analysis of the Balance of Payments in a Model of Accumulation," in Frenkel and H. G. Johnson, eds., *The Monetary Approach to the Balance of Payments*, London 1975.

——, (1975b) "Adjustment Mechanisms and the Monetary Approach to the Balance of Payments: A Doctrinal Perspective," in E. Classen and P. Salin, eds., *Recent Issues in International Monetary Economics*, Amsterdam 1975.

—— and S. Fischer, "International Capital Movements Along Balanced Growth Paths: Comments and Extensions," *Econ. Record*, June 1972, *48*, 266–71.

—— and H. G. Johnson, *The Monetary Approach to the Balance of Payments*, London 1975.

—— and C. A. Rodriguez, (1975a) "Wealth Effects and the Dynamics of Inflation," *J. Money, Credit, Banking*, May 1975, *7*, 259–68.

—— and ——, (1975b) "Inflation and Capital Mobility Under Flexible Exchange

Rates," unpublished paper, Univ. Chicago 1975.

J. P. Gould, "Adjustment Costs in the Theory of Investment of the Firm," *Rev. Fcon. Stud.*, Jan. 1968, *35*, 47–55.

H. G. Johnson, "The Monetary Approach to Balance of Payments Theory," *J. Finance. Quant. Anal.*, Mar. 1972, *7*, 1555–72.

———, "The Monetary Theory of Balance of Payments Policies," in J. A. Frenkel and H. G. Johnson, eds., *The Monetary Approach to the Balance of Payments*, London 1975.

R. Komiya, "Economic Growth and the Balance of Payments," *J. Polit. Econ.*, Feb. 1969, *77*, 35–48.

D. E. Logue and T. D. Willett, "The Effects of Exchange Rate Adjustment on International Investment," unpublished paper, U.S. Treasury Department, 1974.

R. I. McKinnon, "Portfolio Balance and International Payments Adjustment," in R. A. Mundell and A. K. Swoboda, eds., *Monetary Problems of the International Economy*, Chicago 1968.

L. A. Metzler, "Wealth Saving and the Rate of Interest," *J. Polit. Econ.*, Apr. 1951, *59*, 93–116.

R. A. Mundell, *International Economics*, New York 1968.

———, *Monetary Theory*, Pacific Palisades 1971.

M. Mussa, "Tariffs and the Balance of Payments: A Monetary Approach," in J. A. Frenkel and H. G. Johnson, eds., *The Monetary Approach to the Balance of Payments*, London 1975.

J. Myhrman, "Balance of Payments Adjustment and Portfolio Theory: A Survey," in E. Classen and P. Salin, eds., *Recent Issues in International Monetary Economics*, Amsterdam 1975.

D. D. Purvis, "More on Growth and the Balance of Payments," *Can. J. Econ.*, Nov. 1972, *5*, 531–40.

H. Uzawa, "Time Preference and the Penrose Effect in a Two Class Model of Economic Growth," *J. Polit. Econ.*, July/Aug. 1969, *77*, 628–52.

Profit Maximization and Alternative Theories: A Dynamic Reconciliation

By Robert E. Wong*

For a number of decades now the classical concept of the profit-maximizing firm has come under attack. The first major assault may have been the findings of the study by Robert Hall and Charles Hitch in 1939. Their results indicated little support for the use of marginal principles in the real world of business. Their conclusions questioned the validity of the profit-maximizing model of the firm. Since that time a number of alternative theories have emerged with regard to firm objectives and behavior. Probably the most frequently discussed alternative is the sales maximization objective proposed by William Baumol. Baumol's observations as a consultant to business indicated that firms will maximize revenues or sales subject to a constraint of some satisfactory level of profits. J. R. Wilson considers the possibility of dividend maximization by the stockholder controlled firms and undistributed profits maximization by management controlled firms. Armen Alchian and William Allen consider the concept of wealth maximization which essentially corresponds to the maximization of the present value of discounted net income, net receipts, or net cash flow considered by George Stigler, Dale Jorgenson, and Kenneth Arrow. These latter concepts are based upon the streams of returns to the entrepreneur. There has also been a development of a behavioral theory of the firm by Richard Cyert and James March in which individual managerial goals may exist and are not consistent with the organizational goals of profit maximization.

The purpose of this paper is to show that profit maximization and the alternative referenced theories with the exception of the behavioral theory can be reconciled with a dynamic analysis of a firm. The different types of maximizing behavior referenced can be shown to be a part of the expected behavior of the firm operating to maximize the present value of its dividend stream. To perform the reconciliation, a modified Jorgenson dynamic model will be employed. It should be noted that the reconciliation of various maximization concepts with a dynamic analysis was implied in the analysis by Gerald Nordquist.

I. Objective of the Firm

The owners of the corporate form of business are the stockholders. A successful corporation generally will never liquidate its assets, and therefore the stockholder cannot expect that his share of the net worth of the firm will ever be distributed. Thus, the value of his stock is based upon dividends which he receives and expects to receive. In theory the stock value is the discounted stream of dividends of the corporation through an infinite time period. The stockholders would want to maximize their stock value and should direct the activities of the firm's management to do this. We may write the objective as

$$V = \int_0^\infty De^{-rt}dt \tag{1}$$

where D = dividends, r = discount rate, and V = present stock value. The dividend paid to the stockholders is distributed out of net receipts which are defined here as the difference of revenues and expenses in production over a unit of time. The expenses consist only of wages to labor.

That portion of net receipts not distributed to dividends is used for investment in capital. Capital is assumed to be depreciating at a constant rate. The rate of change of capital can be written as

---

* Economist, department of savings and loan, State of California and senior lecturer, University of Southern California. I wish to thank Michael Intriligator, William Barger, and a referee for their helpful comments.

$$(2)\qquad \frac{dK(t)}{dt} = I - \delta K(t)$$

where K = capital, I = investment, and δ = rate of capital depreciation. The expression for the investment can be rewritten as

$$(3)\qquad I = pQ(K(t), L) - wL - D$$

where p = price of output, Q = output based upon neoclassical production function, L = labor, and w = wage rate. This equation for investment assumes that capital financing will be made from internal funds only. It will also be assumed for our analysis that prices and wages are constant. That is, the markets for the firm's product and for labor are perfectly competitive. The output function will be convex as required under the neoclassical assumptions. Furthermore,

$$(4)\qquad \begin{aligned} &0 \leq L \\ &0 \leq D \leq pQ(K(t), L) - wL \end{aligned}$$

which indicates that dividends can never exceed the net receipts, i.e., revenues minus the cost of labor.

II. Optimal Conditions

The above model is essentially the Jorgenson model for the dynamics of the firm somewhat modified. The only material modification is in the limitation on investment (3). Jorgenson assumed only that investment be nonnegative. The mathematical solution for the above problem involves the use of the Maximum Principle of Pontryagin.[1] The firm has two control variables at its disposal, dividend distribution D and the employment of labor L. These control variables are manipulated by management in order to maximize the stockholder's present value of stock. The Hamiltonian is written as[2]

$$(5)\qquad \begin{aligned} H(K, L, D, t) = {} & De^{-rt} + q(t)e^{-rt}(pQ(K(t), L) \\ & - wL - D - \delta K(t)) \end{aligned}$$

where the auxiliary variable q is interpreted as the value of a unit of investment in terms of dividends at time t. To obtain the optimal strategies for dividends and labor, the Hamiltonian is to be maximized at each instant of time with respect to dividend distribution D and labor employment L, subject to the inequalities (4). Let us first examine the characteristics of the behavior of the firm with respect to the optimal dividend distribution and holding labor constant. Since the Hamiltonian is linear in D, its maximization provides three possible optimal alternatives for dividend distribution. These are

$$(6)\qquad D = \begin{Bmatrix} 0 \\ pQ(K, L) - wL - \delta K \\ pQ(K, L) - wL \end{Bmatrix} \quad \text{if } q \begin{cases} > 1 \\ = 1 \\ < 1 \end{cases}$$

where

$$(7)\qquad \frac{dq}{dt} = -q(t)\left[p\frac{\partial Q(K, L)}{\partial K} - (\delta + r)\right] - v(t)p\frac{\partial Q(K, L)}{\partial K}$$

$$v(t) = \begin{Bmatrix} 0 \\ 1 - q \end{Bmatrix} \qquad \text{if } q = \begin{cases} \geq 1 \\ < 1 \end{cases}$$

and

$$\frac{\partial^2 Q(K, L)}{\partial K^2} \leq 0 \qquad \text{if } q = 1$$

This latter inequality is satisfied by the assumed convexity of the production function. For a given value of labor, initial value of capital and q, the growth path of the firm in time is uniquely determined by equations (2), (3), (6), and (7). These paths can be illustrated with the aid of Figures 1 and 2. Figure 1 depicts graphically two components of equations (2) and (3), that is, revenues pQ and costs $wL + \delta K$ as a function of the level of capital. Note that the depreciation of capital is part of the total costs of production. The point of maximum profits is shown to occur at K_m where the marginal revenue with respect to capital equals the slope of the cost function δ. Figure 2 is the phase plane for q and K derivable using Figure 1 and

---

[1] For the elements of this mathematical technique, refer to the text by Michael Intriligator.

[2] For convenience, the (t) will be omitted from q and K in the rest of this paper.

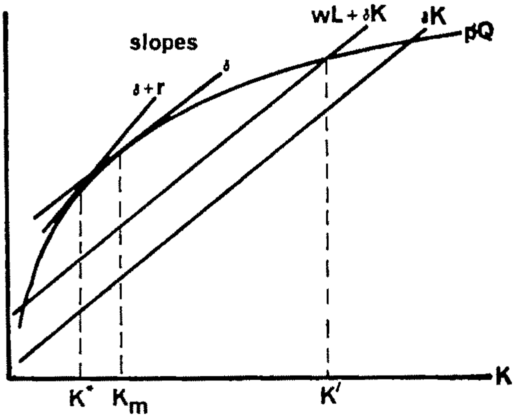


FIGURE 1. REVENUE AND COST FUNCTIONS

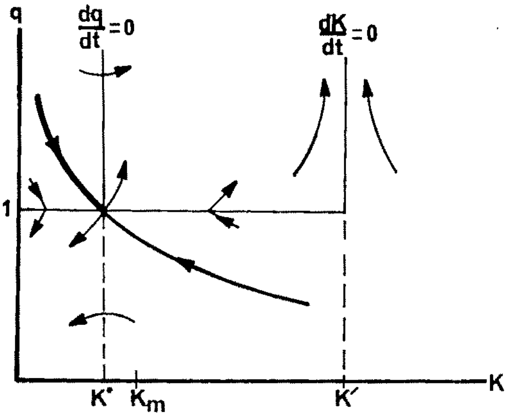


FIGURE 2. PHASE PLANE OF q AND K OPTIMAL DYNAMIC PATHS

equations (6) and (7). The trajectories of Figure 2 indicate the various optimal paths of the firm depending upon initial values of q and K. Alternatively, a specific optimal path for the firm would be determined by specifying the initial value of capital, final value of capital, and the time period.

Of particular interest is the case where the time horizon is infinite, the initial capital $K(0)$ is less than K^* shown in Figure 2, and the final level of capital is allowed to seek the optimal level. Due to this latter specification, the transversality conditions require that $q(t)e^{-rt}$ approach zero as time becomes infinite. The optimal path which satisfies these specifications and the transversality condition is shown by the bold trajectory passing through ($q=1$, K^*). The optimal path in this case would consist of two parts. (See equation (6) for $q>1$ and $q=1$.)

(i) The firm pays no dividends and allocates all net receipts for new capital.
(ii) When the level of capital K^* is reached then the only funds allocated for capital are those needed to replace depreciation. All operating profits are paid to stockholders in the form of dividends.

This trajectory has the turnpike properties found in the Optimal Growth Theory literature such as formulated by David Cass. Path (ii) could be referred to as the "dividend" turnpike and the firm could be said to be operating at its "golden stage" of production.

During the golden stage of production, dK/dt and dq/dt equal zero. That is, the level of capital is constant and the value of a unit of capital in terms of dividends is constant and equal to one. This latter condition requires

$$\frac{\partial Q(K, L)}{\partial K} = \frac{\delta + r}{p} \tag{8}$$

as derived by equation (7). This condition is a marginal productivity condition which is similar to the myopic policy derived by Arrow. The unit cost of capital in production per unit time is the rate of depreciation and the discount rate. The modification over the profit-maximizing condition consists of the discount term.

We have examined the characteristics of the firm's growth under the assumption that labor is fixed. Let us now consider the optimal path of the firm with both capital and labor as variables. Maximizing the Hamiltonian with respect to L requires that

$$\begin{gathered} \frac{\partial Q(K, L)}{\partial L} = \frac{w}{p} \\ \frac{\partial^2 Q(K, L)}{\partial L^2} \leq 0 \end{gathered} \tag{9}$$

hold true at each instant of time. That is, the level of labor employed must be such that the marginal productivity of labor equals the ratio of the unit cost of labor and the price. This condition is the same rule derived in the classical analysis of the profit maximizing firm. The second condition in (9) is a requirement of convexity of the production function relative to labor. The neoclassical assumptions of the model provide satisfaction of this condition.

Thus, the firm will put forth strategies for dividends and labor which will simultaneously satisfy conditions (i) and (9) or (ii), (8), and (9). The combination of (8) and (9) during the golden stage of production is a modified version of the marginal productivity conditions for the classical profit maximizing firm.[3] If the discount rate is zero then the golden stage conditions would be the same as the profit maximizing marginal productivity conditions. The firm in the golden stage is operating at approximately the profit-maximizing state.

An observer viewing the firm in phase (i) might interpret the firm as maximizing sales, growth, revenues, or undistributed profits. In this phase, the firm is expanding capital as rapidly as possible by withholding the distribution of dividends. Thus, the firm's growth and output as well as sales and revenues appear to be increasing as rapidly as possible. When the firm is operating in phase (ii), the observer might interpret the firm as maximizing profits or dividends. Thus, the various alternative theories of the firm noted at the beginning of this paper would all appear to be encompassed as a characteristic exhibited by wealth or stock value maximizing firms sometime during their existence.

III. Stockholder Capital Gains and Income

The results are consistent with general observation in terms of stock prices. Investors generally refer to two categories of stocks for investment, growth stocks and income stocks. Growth stocks are associated with firms which are experiencing a significant degree of expansion. These firms are characterized by little if any dividends and increasing earnings. Investors purchase growth stocks for their appreciation in value, that is, capital gains. These characteristics of growth firms are consistent with the characteristic found in the optimal firm's phase (i). That is, no dividends, increasing earnings, and appreciating value. This latter property is found by the examination of the stock value or objective function at two different time points t_0 and t_1 within phase (i). Let t^* be the time at which dividends are initially paid and phase (ii) begins. Then $t_0 < t_1 < t^*$. Since the dividend is zero in phase (i),

$$V(t_1) = \int_{t_1}^{\infty} D(t)e^{-r(t-t_1)}dt$$

$$> \int_{t_0}^{\infty} D(t)e^{-r(t-t_0)}dt = V(t_0)$$

The stock value at t_1 is greater than at time t_0 since t_1 is closer to the time at which dividends will be paid. That is, due to the discount on the dividend streams, the value at future time points during phase (i) will be greater than at current time points. This can be interpreted as value appreciation or capital gains based upon expectation of dividend payments which corresponds with observation of investor behavior and stock market prices.

Income stocks are characterized by high dividends with relatively stable earnings and relatively stable stock prices. This is consistent with phase (ii) of the optimal firm. In phase (ii), it was found that dividends and profits are constant. Since the dividends paid are constant, then for time points t_2 and t_3 where $t^* < t_2 < t_3$, $D =$ constant; we have

$$\int_{t^*}^{\infty} D(t)e^{-r(t-t^*)}dt = \int_{t_2}^{\infty} D(t)e^{-r(t-t_2)}dt$$

$$= \int_{t_3}^{\infty} D(t)e^{-r(t-t_3)}dt = \frac{D}{r}$$

---

[3] Using the classical calculus of variations, Jorgenson derived some marginal productivity conditions which differed only slightly from equations (8) and (9). However, phase (i) was not obtained by Jorgenson. It should be noted that phase (i) is not derivable with the classical machinery unless the artifice by F. A. Valentine is applied. Bounds were not placed on investment in the positive direction in the Jorgenson model.

or $V(t^*) = V(t_2) = V(t_3)$. That is, the stock value during phase (ii) at any arbitrary time point is constant. This, then, would be consistent with stable stock prices of income stocks.

IV. Summary

The above analysis has examined the optimal dynamic path of the corporate form of the firm. The owners of the corporation are the stockholders whose objective is to maximize the value of their stock. This value is defined as the present value of the discounted stream of dividends over an infinite time. The model is neoclassical. Assuming that the firm is a young firm with a limited amount of initial capital, its future behavior to be optimal would consist of two phases.

(i) *The Growth Maximization Phase* is the phase in which there is investment of all operating profit to provide new capital. No dividends are paid to the stockholders. This maximum rate of capital accumulation continues until the level of capital reaches that required for the golden stage of production. Since the firm is increasing its capital and its output as rapidly as possible, it appears to be maximizing its sales or revenues.

(ii) *The Golden Stage or the Quasi-Profit-Maximizing Phase* is the phase in which the firm behaves in a manner similar to that of the classical profit maximizing firm. That is, the firm will operate near the marginal productivity conditions. The firm will maintain the golden stage level of capital and labor employment. During this phase all operating profits are paid to the stockholders in the form of dividends. Investment is limited to replacement capital.

In effect, the firm is attempting to reach the golden stage of production as soon as possible through its growth maximization phase. Once the golden stage is reached, it is maintained.

The alternative notions to profit maximization such as revenue or sales or growth maximization, dividend maximization, undistributed profits maximization, wealth maximization, and present value maximization would not appear to be unexpected behavior of a corporation whose objective is to maximize stock value. Wealth maximization is equivalent to the objective of stock value maximization. The firm is behaving as if it were maximizing revenues, sales, or growth during phase (i). The firm is producing as much as possible by investing all operating profits in new capital which in turn will increase output.

When the firm has reached the golden stage of production then its behavior could be classified as quasi-profit maximizing or quasi-dividend maximizing. During this phase the firm is distributing all operating profits to the stockholders as dividends. The profits attained are near the level which would be dictated by the classical profit-maximizing case. The difference is due to the rate of discount which favors current dividends over future dividends. As the rate of discount approaches zero, the profits of phase (ii) will approach that of the profit-maximizing case.

The analysis has attempted to reconcile various notions regarding the behavior of the firm. Under the assumption of a neoclassical model for the dynamic firm and the objective of stock value maximization, the reconciliation would appear to be successful.

REFERENCES

A. A. Alchian and W. R. Allen, *Exchange and Production: Theory in Use,* Belmont 1969.

J. K. Arrow, "Optimal Capital Policy with Irreversible Investment," in J. N. Wolfe, ed., *Value, Capital and Growth,* Chicago 1968.

W. Baumol, *Business Behavior, Value and Growth,* New York 1959.

D. Cass, "Optimum Growth in an Aggregative Model of Capital Accumulation," *Rev. Econ. Stud.,* July 1965, *32,* 233–40.

R. M. Cyert and J. G. March, *A Behavioral Theory of the Firm,* Englewood Cliffs 1963.

R. L. Hall and C. J. Hitch, "Price Theory and Business Behavior," *Oxford Econ. Pap.,* May 1939, *2,* 12–45.

M. D. Intriligator, *Mathematical Optimization and Economic Theory,* Englewood Cliffs 1971.

D. W. Jorgenson, "The Theory of Investment Behavior," in R. Ferber, ed., *Determinants of Investment Behavior,* New York 1967, 129–55.

G. L. Nordquist, "The Breakup of the Maximum Principle," *Quart. Rev. Econ. Bus.,* Fall 1965, *5,* 33–46.

G. L. Stigler, *Theory of Price,* New York 1952.

F. A. Valentine, "The Problem of Lagrange with Differential Inequalities as Added Side Conditions," *Contributions to the Calculus of Variations, 1933–37,* Chicago 1937.

J. R. Wilson, "Maximization and Business Behavior," *Econ. Rec.,* May 1952, *28,* 29ff.

The Limits to Growth in a Neoclassical World

By MARTIN J. BECKMANN*

If some resources are exhaustible but essential to production, how should their use be allocated over time? In the presence of technical change, how much resource use is consistent with a stationary state or positive growth?

Questions like these are examined in the Club of Rome study on the limits to growth. Unfortunately, the treatment of preferences, of production technology, and of the economy appears unconventional and implausible to many economists. The following is an attempt to reformulate the allocation problem for exhaustible resources in neoclassical terms.

Preferences are described by a logarithmic utility function and future utility is discounted by a constant discount factor. Production of goods and control of bads (pollution) is described by Cobb-Douglas production functions with exogenous technical progress at constant rates. In addition to producible inputs, net inputs of exhaustible resources are needed in the production of goods. Consumption of leisure is equal to the stock of persons minus the person inputs into production, and in this way labor can be treated as any other consumption good. Under these assumptions which may be more palatable to economic theorists than those of "World Dynamics," the optimal intertemporal allocation of stocks of producible goods and of the given known stocks of exhaustible resources will be investigated.

Our objective is to maximize the present value of the stream of utility generated by consumption (flows) and by the levels of certain variables descriptive of the quality of the environment. The utility function is specified to be log-linear. Such a utility function may be interpreted in the aggregate or in per capita terms without change in the mathematics. It is chosen here because it possesses all essential features of a representative utility function and permits an explicit solution of the problem.

In a given period t let consumption of good i be c_i $i \in s_1$ and let the level of the quality variable j be $q_j j \in s_2$. The utility generated in one period is then

$$(1) \qquad u_t = \sum_{i \in s_1} a_i \log c_i + \sum_{j \in s_2} a_j \log q_j$$

where the coefficients a_i and a_j express the importance of that good or quality variable for well-being. Our objective function is then

$$\sum_{t=0}^{\infty} \rho^t u_t$$

where ρ, $0 < \rho < 1$ is the discount factor, a measure of "impatience." Consider now the technology of production. This is assumed to be Cobb-Douglas and to involve quantities X_{ik} of producible products and quantities s_{hk} of exhaustible resources. We also admit technical change at the rate μ_k. The production function is then

$$(2) \qquad y_k(t+1) = g_k e^{\mu_k t} \prod_i X_{ik}^{\gamma_{ik}}(t) \prod_h s_{hk}^{\delta_{hk}}(t)$$

In addition to producible and exhaustible resources, there may be other resources that are fixed (for example, land). Therefore, the degree of homogeneity may be less than or equal to one.

$$(3) \qquad \sum_i \gamma_{ik} + \sum_h \delta_{hk} \leqq 1$$

The level of quality variable q_j is determined by its previous level, by production activities generating pollutants (described by production inputs) and by control activities introduced to improve the level of this quality variable. The production function for quality variable j is then

$$(4) \qquad q_j(t+1) = g_j e^{\mu_j t} \prod_i X_{ij}^{\nu_{ij}}(t) \prod_h s_{hj}^{\delta_{hj}}(t) \cdot \prod_{ih} X_{ih}^{-\nu_{ihj}}(t) \prod_l q_l^{\epsilon_l}(t)$$

---

[*] Brown University and Institut für Angewandte Mathematik, Technische Hochschule München.

The exponents γ_{ij} and δ_{hj} refer as before to productive inputs into the activity that is to control or improve the quality variable, whereas the exponent ν_{ihj} refers to the generation of pollutants by other production activities.

The optimal production program is best described in terms of the "principle of optimality" of an appropriate dynamic program. Let V_n be the present value of an optimal program of production, consumption, and environmental control extended over n periods. This value depends on the state of the system as described by suitable state variables. These are: the current stocks y_i of produced commodities available for consumption or as inputs into production; the stocks z_h of exhaustible resources available for current or future input into production; the levels of the environmental quality variables q_j; and finally time t itself as an index of the state of technology.

$$(5) \quad V_n = V_n(y_1, \ldots, y_k, \ldots; z_1, \ldots, z_h, \ldots; q_1, \ldots, q_j, \ldots; \mathrm{t})$$

The principle of optimality determines the value function V_n recursively through period by period optimization. The decision to be made in every period is how much of the available quantity y_i of each producible commodity to allocate to current production of every product k, to be denoted by X_{ik}. Consumption is then

$$(6) \qquad c_i = y_i - \sum_k X_{ik}$$

Next it must be decided how much of the available stocks z_h of exhaustible resources h should be used in the production of goods k, s_{hk}. Similarly amounts X_{ij}, s_{kj} must be allocated to the control activities for environmental quality variables q_j. In order to determine the value function, let $V_0=0$. Optimization in a single period requires then that all stocks of products k be consumed.

$$(7) \quad V_i = \sum_i a_i \log y_i + \sum_j a_j \log q_j$$

Now let V_{n-1} be known. The state variables in V_{n-1} depend on the decision variables x_{ik}, s_{hk} in the current period through the production functions (2) and (4). Moreover,

$$(8) \qquad z_h(\mathrm{t}+1) = z_h(\mathrm{t}) - \sum_i s_{hi}$$

The general format of the principle of optimality is

$$(9) \quad V_n(y, z, q, \mathrm{t}) = \max_{x_{ih}, s_{ih}} u\left(y_i - \sum_h x_{ihj} \ldots\right) + \rho V_{n-1}\left[y_k(x_{ih}, s_{ih}), z_h - \sum_i s_{hi}, q_j(\ldots x_{ih} \ldots x_{ij}, s_{hj}, q_t, \mathrm{t})\right]$$

With log-linear utility and production functions it can be shown that the value function is also log-linear

$$(10) \quad V_n = b_{n0} + \beta\mathrm{t} + \sum_k b_{nk} \log y_k + \sum_j b_{nj} \log q_j + \sum_h b_{nh} \log z_h$$

and is linear with respect to time. This is demonstrated for a three-variable model below. In this proof it is also shown that the allocations x_{ik} and s_{hk} are strictly proportional to the available stocks:

$$(11) \qquad x_{ik} = \frac{\alpha_{ik}}{1 + \sum_k \alpha_{ik}} y_i$$

$$(12) \qquad s_{hk} = \frac{\lambda_{hk}}{1 + \sum_k \lambda_{hk}} z_h$$

Let now the planning horizon n be infinite. For the aggregate use of exhaustible resources in every period one obtains a remarkably simple rule: In every period the constant fraction $1-\rho$ of the remaining stock of every exhaustible resource should be used up.

$$(13) \qquad \sum_k s_{hk} = (1-\rho) z_h$$

This fraction is larger the smaller the discount factor ρ for future utility. It is approximately equal to the associated interest rate expressed as a fraction.

$$1-\rho = 1-\frac{1}{1+i} \cong 1-(1-i) = i$$

What are the economic consequences of these decision rules? We will discuss these in terms of a simplified model involving only one producible commodity which is both input and consumption good y, an exhaustible resource z, and one environmental quality variable q. To begin with assume that environmental quality is no problem so that q is also dropped. The utility function is

$$u = a\,log\,(y-x) \tag{14}$$

and the production function

$$y = ge^{\mu t}x^{\gamma}s^{\delta} \qquad \gamma+\delta \leq 1 \tag{15}$$

The dynamic program assumes the form

$$\begin{aligned}b_0+\beta t+b_1\,log\,y+b_2\,log\,z = \\ \operatorname*{Max}_{x,s}\{a\,log\,(y-x)+\rho[b_0+\beta t \\ +b_1\,log\,(ge^{\mu t}X^{\gamma}s^{\delta}) \\ +b_2\,log\,(z-s)]\}\end{aligned} \tag{16}$$

Maximization with respect to x and s requires

$$\frac{-a}{y-x}+\rho b_1\frac{\gamma}{x} = 0 \tag{17}$$

$$\frac{b_1\delta}{s}-\frac{b_2}{z-s} = 0 \tag{18}$$

From (17)

$$x = \frac{\rho b_1\gamma}{a+\rho b_1\gamma}\,y \tag{19}$$

and from (18)

$$s = \frac{b_1\delta}{b_2+b_1\delta}\,z \tag{20}$$

Substitution in (16) yields

$$b_1 = \frac{a}{1-\rho\gamma} \tag{21}$$

and

$$b_2 = \frac{\rho}{1-\rho}\,b_1\delta \tag{22}$$

Therefore,

$$\frac{s}{z} = \frac{b_1\delta}{b_2+b_1\delta} = \frac{1}{1+\dfrac{\rho}{1-\rho}} = 1-\rho \tag{23}$$

$$\frac{x}{y} = \frac{\rho b_1\gamma}{a+\rho b_1\gamma} = \frac{\dfrac{\rho a\gamma}{1-\rho\gamma}}{a+\dfrac{\rho a\gamma}{1-\rho\gamma}} = \rho\gamma \tag{24}$$

Thus

$$z(t) = z(0)\rho^{t} \tag{25}$$

$$s(t) = z(0)(1-\rho)\rho^{t} \tag{26}$$

The production function is then

$$y(t+1) = ge^{\mu t}(\rho\gamma)^{\gamma}y(t)^{\gamma}\rho^{t\delta}(1-\rho)^{\delta}z(0)^{\delta} \tag{27}$$

In logarithmic terms

$$\begin{aligned}log\,y(t+1) = log\,g+\mu t+\gamma\,log\,\rho\gamma \\ +t\delta\,log\,\rho+\delta\,log\,(1-\rho) \\ +\delta\,log\,z(0)+\gamma\,log\,y(t)\end{aligned} \tag{28}$$

or $\eta(t+1) = m+nt+\gamma\eta(t)$

where $\eta = log\,y$. This is solved by

$$\eta(t) = c(t)+d+w(t) \tag{29}$$

where

$$c = \frac{n}{1-\gamma} \tag{30}$$

$$d = \frac{m-\dfrac{n}{1-\gamma}}{1-\gamma} \tag{31}$$

and $w(t)$ is the solution of the homogeneous equation $\eta(t+1) = \gamma\eta(t)$.

$$\text{The system is}\left\{\begin{matrix}\text{growing}\\ \text{stationary}\\ \text{declining}\end{matrix}\right\}\text{according as } n\left\{\begin{matrix}>\\ =\\ <\end{matrix}\right\}0.$$

But

$$n = \mu+\delta\,log\,\rho \tag{32}$$

To avoid decline, technical change in μ must be rapid enough to offset $\delta|log\ \rho|$ which is larger the smaller is ρ. To put it differently, the largest rate of resource use consistent with a stationary state is

$$1 - \beta = 1 - e^{-\mu/\delta} \tag{33}$$

It is larger the larger the rate of technical change μ and the smaller the output elasticity δ of the exhaustible resource. Thus with $\mu = 0.03$ and $\delta = 0.1$ (say) the permissible rate of resource use per period is 26 percent.

Let now one environmental quality variable q be included. The utility function is then

$$u = a_1\, log\,(y - X_1 - X_2) + a_2\, log\, q \tag{34}$$

and the production functions are

$$y = g_1 e^{\mu_1 t} X_1^{\gamma_1} s_1^{\delta_1} \qquad \gamma_1 + \delta_1 \leq 1 \tag{35}$$

$$q' = g_2 e^{\mu_2 t} X_1^{-\nu} X_2^{\gamma_2} s_2^{\delta_2} q^{\epsilon} \tag{36}$$

where q' denotes the quality level of the next period. The principle of optimality of the dynamic program is

$$\begin{aligned} &b_0 + \beta t + b_1\, log\, y + b_2\, log\, q + b_3\, log\, z \\ &= \operatorname*{Max}_{x_1, x_2, s_1, s_2} \{a_1\, log\,(y - X_1 - X_2) + a_2\, log\, q \\ &\quad + \rho[b_0 + \beta t + b_1\, log\,(g_1 e^{\mu_1 t} X_1^{\gamma_1} s_1^{\delta_1}) \\ &\quad + b_2\, log\,(g_2 e^{\mu_2 t} X_1^{-\nu} X_2^{\gamma_2} s_2^{\delta_2} q^{\epsilon}) \\ &\quad + b_3\, log\,(z - s_1 - s_2)]\} \end{aligned} \tag{37}$$

Maximization with respect to X_1, X_2, yields

$$\frac{-a_1}{y - X_1 - X_2} + \frac{\rho b_1 \gamma_1}{X_1} - \frac{\rho b_2 \nu}{X_1} = 0 \tag{38}$$

$$\frac{-a_1}{y - X_1 - X_2} + \frac{\rho b_2 \gamma_2}{X_2} = 0 \tag{39}$$

from which

$$X_1 = \frac{b_1\gamma_1 - b_2\nu}{\dfrac{a_1}{\rho} + b_1\gamma_1 + b_2\gamma_2 - b_2\nu}\, y \tag{40}$$

$$X_2 = \frac{b_2\gamma_2}{\dfrac{a_1}{\rho} + b_1\gamma_1 + b_2\gamma_2 - b_2\nu}\, y \tag{41}$$

Maximizing with respect to s_1 and s_2 yields

$$\frac{b_1\delta_1}{s_1} - \frac{b_3}{z - s_1 - s_2} = 0$$

$$\frac{b_2\delta_2}{s_2} - \frac{b_3}{z - s_1 - s_2} = 0$$

$$\frac{s_2}{s_1} = \frac{b_2\delta_2}{b_1\delta_1} \tag{42}$$

from which (and equation (13))

$$s_1 = \frac{1 - \rho}{1 + b_2\delta_2/b_1\delta_1}\, z \tag{43}$$

$$s_2 = \frac{1 - \rho}{1 + b_1\delta_1/b_2\delta_2}\, z \tag{44}$$

Substituting in the production function and writing $\eta = log\ y$ as before

$$\begin{aligned} \eta(t + 1) &= log\, g_1 + \mu_1 t \\ &\quad + \gamma_1\, log\, \frac{b_1\gamma_1 - \nu b_2}{\dfrac{a_1}{\rho} + b_1\gamma_1 + b_2\gamma_2 - b_2\nu} \\ &\quad + \gamma_1\eta(t) + \delta_1\, log\, \frac{1}{1 + \dfrac{b_2\delta_2}{b_1\delta_1}} \\ &\quad + \delta_1 t\, log\, \rho \end{aligned} \tag{45}$$

This is again of the form

$$\eta(t + 1) = m + nt + \gamma_1\eta(t)$$

and is solved by

$$\eta(t) = ct + d + \omega(t) \tag{46}$$

where $\omega(t)$ is the solution of the homogeneous equation

$$\eta(t + 1) = \gamma_1\eta(t) \tag{47}$$

which may be disregarded in the long run (as before) since

$$\omega(t) = \gamma_1^{t}\omega(0) \rightarrow 0 \tag{48}$$

since $0 < \gamma_1 < 1$. Now the coefficients in (46) are

$$c = \frac{n}{1-\gamma} \tag{49}$$

$$d = \frac{m - \dfrac{n}{1-\gamma}}{1-\gamma} \tag{50}$$

Again output is $\left\{\begin{matrix}\text{increasing}\\ \text{stationary}\\ \text{declining}\end{matrix}\right\}$ according as

$$n = \mu_1 + \delta_1 \mathit{log}\, \rho \left\{\begin{matrix} > \\ = \\ < \end{matrix}\right\} 0.$$

As an additional requirement for feasibility of long-run production we have that the coefficient of X_1 be positive.

$$b_1\gamma_1 - b_2\nu > 0 \tag{51}$$

which imposes an upper bound on the pollution elasticity ν. Consider next the quality of the environment

$$\mathit{log}\, q(\mathrm{t}) = \kappa(\mathrm{t}) \tag{52}$$

$$\begin{aligned} \kappa(\mathrm{t}+1) &= \mathit{log}\, g_2 + \mu_2\mathrm{t} - \nu\, \mathit{log}\, X_1 \\ &\quad + \gamma_2\, \mathit{log}\, X_2 + \delta_2\, \mathit{log}\, s_2 + \epsilon\kappa(\mathrm{t}) \\ &= \text{constant} + \mu_2\mathrm{t} + \delta_2\mathrm{t}\, \mathit{log}\, \rho \\ &\quad + (\gamma_2 - \nu)[\mu_1 + \delta_1\, \mathit{log}\, \rho]\mathrm{t} + \epsilon\kappa(\mathrm{t}) \end{aligned} \tag{53}$$

by the same argument as for output $\eta(\mathrm{t})$. This is increasing or stationary, whenever

$$\begin{aligned} &\mu_2 + \delta_2\, \mathit{log}\, \rho \\ &\qquad + (\gamma_2 - \nu)[\mu_1 + \delta_1\, \mathit{log}\, \rho] \geqq 0 \end{aligned} \tag{54}$$

or

$$\rho \leqq \exp\,[(\mu_2 + (\gamma_2 - \nu))/(\delta_2 + (\gamma_2 - \nu)\delta_1)] \tag{55}$$

A further possibility is that output increases and the quality of the environment declines (or vice versa) while the utility level increases or is stationary. Now utility is

$$u = \text{constant} + a_1\eta(\mathrm{t}) + a_2\kappa(\mathrm{t}) \tag{56}$$

and this is a linear function of time.

$$\begin{aligned} u(\mathrm{t}) = \text{constant} &+ a_1[\mu_1 + \delta_1\, \mathit{log}\, \rho]\mathrm{t} \\ &+ a_2[\mu_2 + \gamma_2\mu_1 - \nu\mu_1 \\ &\qquad + (\delta_2 + (\gamma_2 - \nu)\delta_1)\, \mathit{log}\, \rho]\mathrm{t} \end{aligned} \tag{57}$$

The condition for utility not to decline is, therefore,

$$\begin{aligned} &a_1\mu_1 + a_2\mu_2 + a_2(\gamma_2 - \nu)\mu_1 \\ &\quad + [a_1\delta_1 + a_2\delta_2 + a_2\gamma_2 - \nu\delta_1]\, \mathit{log}\, \rho \geqq 0 \end{aligned} \tag{58}$$

from which

$$\begin{aligned} &1 - \rho \leq 1 \\ &\quad - \exp\left\{-\frac{a_1\mu_1 + a_2\mu_2 + a_2(\gamma_2 - \nu)\mu_1}{a_1\delta_1 + a_2\delta_2 + a_2(\gamma_2 - \nu)\delta_1}\right\} \end{aligned} \tag{59}$$

In any case positive technical progress is required in order that resource exhaustion should not lead to an eventual inevitable decline of the utility level. The rate of technical change properly weighed with the coefficients of the utility and the elasticities of the production function impose an upper bound on the permissible rates of exhaustion. As an illustration suppose that $a_1 = a_2$, $\mu_1 = \mu_2 = 0.03$. $\delta_1 = \delta_2 = 0.2$, $\gamma_1 = \gamma_2 = 0.8$, $\nu = 0.1$. Then

$$1 - \hat{\rho} \leq 1 - e^{-0.15}$$

$$1 - \hat{\rho} \cong 14 \text{ percent} \tag{60}$$

REFERENCES

T. C. Koopmans, "Some Observations on 'Optimal' Economic Growth and Exhaustible Resources," in H. C. Bos, ed., *Economic Structure and Development, Essays in Honor of Jan Tinbergen,* Amsterdam 1973, 239–55.

D. H. Meadows et al., "The Limits to Growth," report for The Club of Rome's Project on the Predicament of Mankind, New York 1972.

W. D. Nordhaus, "World Dynamics: Measurement Without Data," *Econ. J.,* Dec. 1973, *83,* 1156–83.

R. M. Solow, "The Economics of Resources or the Resources of Economics," *Amer. Econ. Rev. Proc.,* May 1974, *64,* 1–14.

Capital Gains and the Aggregate Consumption Function: Comment

By MICHAEL B. MCELROY AND J. C. POINDEXTER*

A direct causal relationship between capital gains and consumption expenditures was hypothesized by J. M. Keynes in the *General Theory* (pp. 92–94). Only recently, however, have empirical studies of the consumption function attempted to verify this relationship. Making use of his own empirical estimates of accrued capital gains and losses, Kul Bhatia, in a recent article in this *Review*, claims to have done the following:

> We incorporate capital gains into the permanent income framework and test several hypotheses about the effect of capital gains on aggregate consumption. . . . To anticipate the conclusions of the paper, when an appropriate distributed lag function is specified, both accrued and realized capital gains turn out to be significant variables in the aggregate consumption function. [p. 866]

However, Bhatia's statistical results and therefore the ". . . important implications for taxation and other areas of economic analysis and policy" (p. 876) that he draws are called into question when some tenuous and highly restrictive aspects of his specification of the consumption function are amended. In this note we argue that Bhatia's model: (a) includes the same variable, capital gains, twice—once as a separate argument and again in the net worth term; (b) unjustifiably constrains the coefficients on expected gains and initial wealth to be equal; (c) assumes money illusion by failing to eliminate purely inflationary increments from the capital gains series; and (d) incorrectly employs only nominal measures of the variables in the empirical estimation.

In his wealth approach formulation of the consumption function,[1] Bhatia estimates permanent income as the sum of expected disposable labor income (Y_t^w) and expected income from wealth (βW_t^*) where W_t^* is an "expected wealth" variable as discussed below and β is the "normal return" on wealth. He then fits the following function:

$$C_t = kY_t^w + k\beta W_t^*$$

or, letting $W_t^* = W_{t-1} + \omega(L)G_t$,

$$C_t = kY_t^w + k\beta W_{t-1} + k\beta\omega(L)G_t \tag{1}$$

where C_t is consumption in period t, W_{t-1} is wealth at the end of the previous period, $\omega(L)$ is a lag operator, and G_t is the accrued capital gain (or loss) in period t.[2] According to Bhatia, "The term W_t^* . . . is a subjective value concept which differs from the actual market value of wealth inasmuch as it includes *expected* gains $[G_t^e]$ whereas only *actual* gains would be included in W_t. . . . If our specification is valid, the gains' coefficient in [(1)] should be positive and significant" (p. 868).

---

* North Carolina State University. We are grateful to Bhatia for making some of his unpublished data available to us and to T. Dudley Wallace for helpful discussions.

[1] Bhatia distinguishes between an "income" and a "wealth" approach to incorporating capital gains as an independent variable in the aggregate consumption function. We restrict our comments to the latter formulation since his empirical tests using the income approach fail to support his hypothesis that capital gains are a significant determinant of consumer spending. Bhatia contends that the regression results obtained in the wealth approach confirm the hypothesis that both accrued and realized capital gains affect consumption (via wealth), but he gives no theoretical rationale for preferring this specification over the income approach. Even if we did not have reason to doubt the validity of his specification and testing of the wealth approach, we would be skeptical of his assertion, based primarily on a comparison of the Durbin-Watson statistics, that these results appear "much sharper" (p. 875). The Durbin-Watson statistic may not be a particularly appropriate criterion for choosing between these two specifications. In regressions involving a lagged dependent variable (past consumption is a component of the lagged income term), this statistic is biased toward indicating no serial correlation.

[2] We are primarily concerned with the accrued gains-consumption relationship since, as Bhatia notes, two-way causation is likely to be a problem when testing for the role of realized capital gains. There is evidence that individuals may realize their gains *after* the consumption decision has been made. Furthermore, the realization process is simply an asset transformation and while its influence on consumption behavior is certainly not without interest, it is far less basic than the accrued

However, the specification is not valid. The market value of wealth at the end of the preceding period (W_{t-1}) *already embodies* expectations of future income from asset holdings. Even if individuals should extrapolate past capital gains experience in the manner assumed by Bhatia (a distributed lag expectation model), all of their experience through the end of the preceding period ($t-1$) would be reflected in the expectation in period $t-1$ of future gains and, unless asset markets are seriously flawed (an argument which Bhatia does not rely on), those expectations would be fully capitalized and reflected in W_{t-1}. The only additional information on expected future capital gains that would be provided by Bhatia's distributed lag function on capital gains would come from the current period's gain G_t.

This flaw in Bhatia's wealth approach may be seen more clearly by spelling out the components of the wealth term. While W_{t-1} consists of more than just accumulated capital gains (all net capital formation is included), it must include $(G_{t-1}+G_{t-2}+\ldots+G_{t-n}+\ldots)$. Bhatia's expected capital gains function in equation (1) takes the form $(\omega_0 G_t+\omega_1 G_{t-1}+\omega_2 G_{t-2}+\ldots+\omega_n G_{t-n}+\ldots)$. Equation (1) clearly fails to perform the test Bhatia intends for the independent, separable effect of capital gains on consumption. The only term appearing in the expected capital gains variable that does not also appear in the expression for W_{t-1} is $\omega_0 G_t$, weighted *current* capital gains.

Deleting the repetitive lagged terms in his specification yields:

$$C_t = k_0 Y_t^w + k_0\beta W_{t-1} + k_1 G_t \tag{2}$$

which, fitted with Bhatia's data, gives[3]

$$\underset{}{C_t =} \underset{(39.20)}{.780Y_t^w} + \underset{(4.19)}{.052W_{t-1}} + \underset{(0.56)}{.008G_t}$$

Bhatia is right to argue that this equation showing a statistically insignificant role for capital gains does not test for the independent effect of previous capital gains on consumption. However, neither does his equation (1).

A second basic shortcoming in Bhatia's analysis is that he constrains the coefficient ($k\beta$ in equation (1)) on expected capital gains (G_t^e) to be equal to the coefficient on last period's measured wealth. It appears that this restriction accounts for the statistical significance of his capital gains term. This is shown in Table 1, which summarizes the results we obtained using Bhatia's data and lag weights in regressions on equation (1) both with and without the equality constraint. The unconstrained estimates (3) and (4) yield about the same coefficients on the income and wealth terms as the constrained form (1) and (2). But relaxing the equality constraint results in a statistically insignificant coefficient on accrued gains.[4] Hence, Bhatia's results depend upon a critical behavioral hypothesis—that individuals treat all increments to their net worth identically in their consumption decision. This hypothesis is not supported by his data.

The lack of richness of the data does not

---

gains-consumption relationship. In other words, the coefficient on realized gains represents the effect of a particular change in the *composition* of wealth on consumption, whereas the coefficient on accrued gains represents the effect of a change in the *level* of wealth on consumption.

[3] Expected labor income (Y_t^w) was generated using a second-degree Almon lag operator to approximate a fourth-degree polynominal lag. These results are consistent with earlier findings such as those of John Arena. The t-statistics are in parentheses below the coefficients. The R^2 for this regression is .999, the *D.W.* is 1.74, and the *SEE* is 2.32. The sums of squared errors were 44.85 and 43.0, respectively, for equations (1) and (2) in our regressions with Bhatia's data, indicating that no explanatory power is lost by deleting the repetitive terms.

[4] Though we have focused on the accrued gains-consumption relationship, it should be noted that dropping the equality constraint also causes the coefficient on realized gains to become statistically insignificant.

TABLE 1—ESTIMATES OF THE COEFFICIENTS OF BHATIA'S WEALTH APPROACH WITH AND WITHOUT THE EQUALITY CONSTRAINTS
(for accrued capital gains)[a]

	Constant	Y_t^w	W_t^*	W_{t-1}	G_t^e	*SEE*	R^2	*DW*
1)	4.792	.702	.062			2.08	.999	1.96
	(0.54)	(5.22)	(3.27)					
2)		.790	.052			2.02	.999	1.66
		(19.78)	(7.28)					
3)	4.860	.699		.062	.077	2.19	.999	1.93
	(0.52)	(4.89)		(3.12)	(0.90)			
4)		.793		.051	.037	2.11	.999	1.70
		(17.89)		(6.89)	(0.54)			

[a] Regressions 1) and 2) impose the equality constraint on the coefficients of expected capital gains. Regressions 3) and 4) drop this constraint. All variables are in billions of dollars and *t*-statistics are given in parentheses below the coefficients.

permit us to reject the hypothesis that the coefficients on the lagged wealth and expected capital gains terms are equal. However, neither is it capable of refuting the hypothesis that accrued capital gains have *no place* in the consumption function.

Another point that should be noted in evaluating Bhatia's results is that he measures capital gains accruing in period t as the change in net worth ($W_t - W_{t-1}$) not attributable to measured net capital formation (I_t), i.e.,

$$G_t \equiv W_t - W_{t-1} - I_t$$

When changes in the price level occur, the variables on the right-hand side of this identity will be measured in different units, introducing an illusory increment into his measure of accrued gains. There seems to be no a priori rationale for building this money illusion into the consumption function. A more useful measure of accrued capital gains, net of the inflationary increment, would be:

$$G_t' \equiv W_t - \left(\frac{P_t}{P_{t-1}}\right) W_{t-1} - \left(\frac{P_t}{\bar{P}_t}\right) I_t$$

where P_t is an end of period t implicit price deflator and $\bar{P}_t$ is an "average" implicit price deflator for period t.[5] This is still in current dollars but eliminates Bhatia's mismeasurement by converting all the terms on the right-hand side into the same units. Even with an otherwise correctly specified consumption model, a significant coefficient on his G_t term could result from the correlation between this inflationary increment and the price component of the *nominal* consumption measure he employs.[6]

It should also be stressed that all of Bhatia's variables are measured in current rather than constant dollars. His failure to deflate is puzzling indeed since, as is well known, this results in a further element of spurious correlation in the estimated coefficients.

We agree with Bhatia that the role of capital gains in consumption has important policy implications. However, we do not believe that his analysis has added to our knowledge in this area. He argues that earlier studies of the role of capital gains in the consumption function failed to find any significant influence because ". . . capital gains have not been studied in an appropriate theoretical model" and because no ". . . direct estimates of gains have been used in empirical estimation" (p. 866). We have argued that he has not provided an "appropriate" model since his expected capital gains term is not independent of lagged wealth. Moreover, he has confused assumption with empirical finding by building an empirically unverified constraint into the capital gains coefficient. Thus, his policy conclusions for the economic effects of capital gains taxation are based on an *assumed*, not a statistically verified, behavioral relationship. Finally, we contend that Bhatia's estimates of accrued capital gains are inappropriate for the tests he performs and that his use of nominal measures throughout cannot be justified. Empirical verification of the intuitively reasonable hypothesis that real accrued capital gains and losses do affect the consumption decision remains elusive.

---

[5] This adjustment makes a significant difference in the magnitude, alters the year-to-year variation, and occasionally even changes the sign of estimated accrued capital gains. For example, McElroy, using a slightly broader base for estimating capital gains, found that "gross" money gains G from 1946 through 1968 totalled \$1173 billion, whereas "net" money gains G' were \$598 billion.

[6] The simple correlation between P and ΔP is $-.30$.

REFERENCES

J. J. Arena, "Capital Gains and the 'Life Cycle' Hypothesis of Saving," *Amer. Econ. Rev.*, Mar. 1964, *54*, 107–11.

K. B. Bhatia, "Capital Gains and the Aggregate Consumption Function," *Amer. Econ. Rev.*, Dec. 1972, *62*, 866–79.

J. M. Keynes, *The General Theory of Employment, Interest and Money,* New York 1936.

M. B. McElroy, "Capital Gains and the Concept and Measurement of Purchasing Power," in *Proc. Amer. Statist. Assn., Bus. and Econ. Statist. Sec. 1970,* Washington, 132–39.

Capital Gains and the Aggregate Consumption Function: Reply

By KUL B. BHATIA*

Michael McElroy and J. C. Poindexter (M-P hereafter) seek to demolish the theoretical structure and the main findings of my paper in their comment. However, their criticisms are not well founded and are highly conjectural in nature. Their major criticism about the specification of the "wealth approach" is invalid, they do not exactly duplicate the statistical tests used in my paper, and they do not show how the data adjustments proposed by them will affect the final results.

I. Equations in the Wealth Approach

The consumption function in the wealth approach in my paper (equation (6)) is $C_t = kY_t^w + k\beta W_t^*$ where k is the marginal propensity to consume permanent income, "β is the 'normal return' on wealth, and W_t^* is the expected wealth relevant for making consumption decisions in period t . . ."; W_t^* is defined as "the sum of actual wealth at the end of the previous period (W_{t-1}), and a lagged function of accrued capital gains [$W(L)G_t$]" (p. 868). Substitution for W_t^* in (6) yields

$$C_t = kY_t^w + k\beta W_{t-1} + k\beta W(L)G_t$$

which is equation (8) in the paper. M-P contend that there are two basic flaws in this analysis: first, it ". . . includes the same variable, capital gains, twice—once as a separate argument and again in the net worth term," and second, it ". . . unjustifiably constrains the coefficients on expected gains and initial wealth to be equal" (p. 700). . . . "It appears that this restriction accounts for the statistical significance of his capital gains term" (p. 701). Consequently, they delete the "repetitive lagged terms" and present unconstrained estimates in their Table 1. As the following discussion will show, both these points relate to the expected wealth term W_t^* in the consumption function.

A. *The Specification of Expected Wealth* (W_t^*)

The derivation of W_t^* in my paper can be elaborated on as follows:

$$(1) \qquad W_t \equiv W_{t-1} + A_t + G_t$$

$$(2) \qquad W_t^* = W_{t-1} + G_t^e$$

$$(3) \qquad G_t^e = W(L)G_t$$

where W_t and W_t^* are as defined above; A_t and G_t, respectively, are net acquisitions and accrued capital gains during t; and the expected accrued gains during t are denoted by G_t^e which is derived from a distributed lag function of current and past accrued gains. For the empirical results derived in the paper, $W(L)G_t$ was defined by a geometric lag function $(1-\lambda)\ \sum_{i=0}^{t} \lambda^i G_{t-i}$. Using this specification we get:

$$(4) \qquad W_t^* = W_{t-1} + \sum_{i=0}^{t} w_i G_{t-i}$$

where $w_0 = (1-\lambda)$, $w_1 = \lambda(1-\lambda)$,
$w_2 = \lambda^2(1-\lambda)$, etc.

M-P object to equation (4) because W_{t-1} already includes G_{t-1}, G_{t-2}, etc., and these are also present in the expected gains term, G_t^e. Deleting the repetitive lagged terms, they write the consumption function (their equation (2)) as

$$C_t = k_0 Y_t^w + k_0\beta W_{t-1} + k_1 G_t$$

which implies

$$(5) \qquad W_t^* = W_{t-1} + w_0 G_t$$

But why should the repetitive lagged terms be deleted? If we lag equation (1) by one period, and the terms on the right-hand side successively, we derive:

---

* University of Western Ontario. Thanks are due to my colleagues C. Scott Clark and E. J. D. Ketchum for helpful discussions.

$$(1') \quad W_{t-1} = W_0 + \sum_{i=1}^{t} G_{t-i} + \sum_{i=1}^{t} A_{t-i}$$

where W_0 is initial wealth. In other words, the market value of wealth at the end of year $t-1$ is the sum of initial wealth (W_0), all gains which have accrued until then ($\sum G_{t-i}$), and net acquisitions ($\sum A_{t-i}$). Substituting (1′) into (4) and regrouping terms gives

$$(6) \quad W_t^* = W_0 + \sum_{i=1}^{t} A_{t-i} + w_0 G_t + \sum_{i=1}^{t} (1 + w_i) G_{t-i}$$

To derive M-P's equation (2), (6) has to be supplanted by (5). This implies that $W_{t-1} = W_0 + \sum_{i=1}^{t} A_{t-1} + \sum_{i=1}^{t} (1+w_i) G_{t-i}$, which violates (1) and (1′) unless $w_i (i=1, 2, \ldots t) = 0$. Alternatively, if $w_i (i=1, 2, \ldots t)$ are indeed zero, equation (3) is contradicted. Either contradiction has rather serious implications. If equation (3) holds, W_{t-1} depends directly on w_i which are the weights that minimize the residual sum of squares and yield the maximum likelihood estimates of the consumption function (see fn. 17 in the original paper). On the other hand, if $w_i (i=1, 2, \ldots)$ are zero, it implies that expectations about future gains are based on accrued gains for the current period only. Both (1) and (1′) are well established and widely used accounting definitions of market value of wealth, and the expectations mechanism in equation (3) is an important element of the literature on the permanent income hypothesis and distributed lags. Deleting the repetitive lagged terms involves a contradiction which M-P do not seem to recognize. Their justification for this deletion is vague and unconvincing. They do not appear to reject the expectations mechanism in their comment, but on page 701 they state, "The market value of wealth at the end of the preceding period (W_{t-1}) *already embodies* expectations of future income from asset holdings." Might we assume, therefore, that M-P are offering a new definition of market value of wealth which includes regression parameters?

Another objection raised by M-P is implied by their remark that if lagged gains are included in both W_{t-1} and G_t^e, one cannot perform a test "for the independent, separable effect of capital gains on consumption" (p. 701). This is a statistical argument, not a theoretical one, and its force depends on whether the constrained estimation used in my paper is justified or not. Before dealing with this issue, however, it should be emphasized that, as the discussion here clearly shows, the wealth equations in my paper are correctly specified and consistent with the theoretical framework set out there. M-P's equation (2) and equations (3) and (4) in their Table 1 do not follow from the model presented in my paper.

B. *Constrained vs. Unconstrained Estimation*

For deriving the empirical results in my paper the coefficients of W_{t-1} and G_t^e are constrained to be equal because W_t^* is defined by (2). The constraint thus follows directly from the theoretical specification of the consumption function and implies that expected income from wealth (βW_t^*), irrespective of the components of W_t^*, is the relevant variable (besides other permanent income) for consumption decisions. Even as a practical matter, it is reasonable to assume that changes in market value of net worth, whether caused by variations in asset prices or net acquisitions, will have the same effect on consumption. A case could perhaps be made that people treat, say, corporate stock holdings differently from real estate holdings, but to treat net acquisitions and *expected* capital gains on the same asset differently does not seem very plausible.

In this context, it is easy to speculate but difficult to test whether the constraint accounts for the statistical significance of the capital gains term. The test conducted by M-P is incorrect because it is inconsistent with the theoretical model in my paper. To perform the test correctly, let us define W_t^i as the sum of initial wealth and all past net acquisitions.

$$W_t' = W_0 + \sum A_{t-i} \tag{7}$$

$$W_t = W_t' + \sum_{i=0}^{t} G_{t-i} \tag{8}$$

W_t' thus is different from W_t because it does not include any capital gains. If wealth and expected gains have different coefficients, the consumption function can be rewritten as

$$C_t = k_0 Y_t^{\omega} + k_1 W_{t-1} + k_2 G_t^{e} \tag{9}$$

Substituting from (8) for W_{t-1} and using the specification $G_t^e = \sum w_i G_{t-i}$ yields

$$C_t = k_0 Y_t^{\omega} + k_1 W_{t-1}' + k_1 \sum_{i=1}^{t} G_{t-i} + k_2 \sum w_i G_{t-i}$$

or

$$C_t = k_0 Y_t^{\omega} + k_1 W_{t-1}' + k_2 w_0 G_t + \sum_{i=1}^{t} (k_1 + w_i k_2) G_{t-i} \tag{10}$$

The correct unconstrained form of the consumption function thus is provided by (10). It calls for complicated, non-linear estimation which definitely lies outside the scope of this brief reply. In their Table 1, M-P report two regression estimates of equation (9), and not of equation (10). That clearly is not a test of the constraint specified in the theoretical model of my paper.[1]

II. The Measurement of Accrued Gains

M-P also criticize my use of nominal rather than real magnitudes, and the measure of capital gains adopted in my paper. As a general proposition, I agree that wherever possible aggregate consumption functions should be estimated not only in real but also in per capita terms. My choice of nominal variables, therefore, is based mainly on empirical considerations.

The task of computing even ***nominal*** gains is beset with enormous data difficulties caused by conflicting series on market value, missing data, incomplete information in several areas, and highly deficient documentation of published series. Corporate stock accounted for more than 60 percent of all accrued gains between 1948 and 1964, the time period covered by my study. As discussed at length in my 1973 paper, several adjustments were necessary to derive consistent estimates of nominal gains accruing on corporate stock. Estimating real gains, as M-P recognize, is not a matter of simply deflating nominal gains by a price index. It requires deflation of the asset holdings at the beginning of a year, at the end of the year, and also of net acquisitions during the year. Now, ignoring for the moment the problem of unsatisfactory deflators, there are no adequate data on net acquisitions of corporate stock.[2] Therefore, although some indirect procedures are possible, they will further complicate, and perhaps also introduce additional errors, in estimating ***real*** capital gains. The use of nominal variables throughout is the obvious next-best solution.

Another difficulty caused by using real gains is that with changes in price level, gains or losses will accrue on all assets whose face values are fixed in nominal terms. Thus, in an inflationary situation, losses will accrue on people's cash balances, demand deposits, etc., and gains will accrue on the net borrowing of households from other sectors of the economy. Estimating gains and losses on these additional items will create further problems because of inadequacies in the available data.

Although M-P's representation of my method of estimating gains is only partly accurate,[3] their point, also stated in my paper (p. 870, fn. 9), is correct that no adjustment for changes in price level is made in deriving the

---

[1] A statistical point about the results in Table 1 should also be noted: M-P use the weights for Y_t^{ω} and G_t^e derived in estimating 1) and 2), also in equations 3) and 4). (This is not stated in the comment but has been confirmed by McElroy in a letter to the author.) These weights were estimated in the regression results reported in my paper. Since they were optimally chosen for equations 1) and 2), it is doubtful whether they minimize the residual sum of squares of equations 3) and 4) in Table 1 (see fn. 17 of my paper for details of the search procedure by which the optimal weights were selected). Even on its own ground, therefore, M-P's test is statistically unsound.

[2] For a critical discussion of the data on net acquisitions of corporate stock, see the author (1973), pp. 56–58, and the references cited there.

[3] For my paper, accrued gains on farm assets and

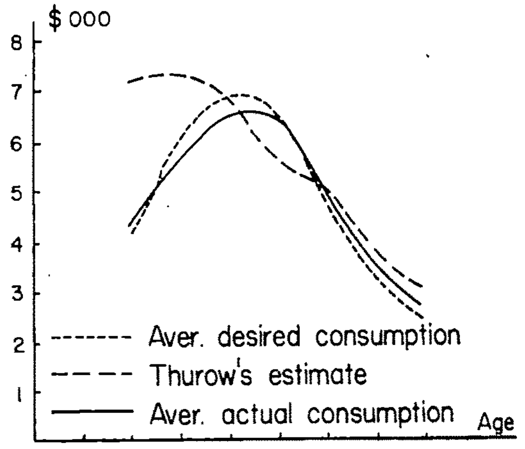


FIGURE 6

Taking the percentile profiles in Figure 3 to approximate income profiles at different levels of wealth, we associate with each a homothetic transform of the average income profile (the coefficient, in each case, being such that the transform coincides with the associated profile at age 50), and the deviations at each age between profiles and transforms are used to obtain the following rough estimates of $r(A)$.[19]

Age:	20	30	40	50	60	70	80
$r(A)$:	−1263	−1595	−963	0	1910	2808	3211

Introducing these estimates together with $p=0.80$ and Census measures of $F(A)$ in relation (29′), we obtain the corrected estimates of $g(A)$ shown as a thick dotted line in Figure 6. *They do reveal an average pattern of desired consumption above actual in the age range 25 to 45, with maximum discrepancy at age 35, but the gap is small in relation to that calculated by Thurow, and it occurs over a later period of the life cycle.*[20] It would appear, in particular, that the very young would continue to opt for the less expensive pleasures of life were the capital market open more widely to them.

E

One disturbing feature of the desired-consumption profile estimated in the previous section is that it exceeds actual consumption over a range (age 25–45) in which income is itself higher than consumption: If the average-income household wishes to consume more in that range, why does it not do so by spending more of its income? The main answer, of course, is that the definition of consumption used so far excludes personal-insurance premiums and gifts or contributions, so that some expenditures of present benefit (utility) to the household are incorporated in the measure of savings. If we adjust the average income profile by subtraction of contributions and insurance at each age, so as to obtain a closer identification of the income-consumption differential with savings, we find that the profile stays just below actual consumption up to age 45, then exceeds it in years beyond. This is perfectly consistent with a pattern of desired consumption in excess of actual in the 25–45 age interval and below actual in later years.

On the other hand, the expected near-equality of desired and actual consumption in zero-saving situations cannot hold irrespective of whether savings are defined to include or exclude certain items. Thurow's liberal offer of a choice between two estimates obtained by applying the same methodology to two definitions of savings weakens confidence in the validity of his sampling.

---

[19] With the a priori specification: $r(50)=0$, the relative wealth W^i under income profile i is obtained as $y^i_{50}/\bar{y}_{50}$, where y^i_A is profile i income at age A and $\bar{y}_A$ is average income at age A. If the profiles behave in accordance with (19′) without error, we have:

$$y^i_A = W^i[\bar{y}_A + r(A)] - r(A)$$
$$r(A) = \frac{y^i_A - W^i\bar{y}_A}{W^i - 1}$$

The above expression is used to compute three point estimates of $r(A)$—one for each of the three profiles—at each age of observation. The figures shown for each age above are simple averages of the three point estimates.

[20] The plot of actual consumption is that of average consumption under Thurow's "first" definition—exclusive of personal transfers and insurance. The diagram also shows Thurow's own estimate of $g(A)$ under the corresponding definition. Note that the identification of average consumption as the consumption of a household with average income can only be accepted as an approximation.

(27) and (28) becomes $P=1/b$, and the weight of identification falls entirely on the biasing term $\psi(W)$.[16] This should not have prevented Motley and Morley from deriving P by reference to the data at hand, as was done in Section IVA. However, the temptation was strong to utilize external evidence already processed under similar assumptions, namely, estimates of P derived by Milton Friedman in substantiation of his permanent-income hypothesis.[17] Since there is far from unanimous agreement concerning these estimates, the door is open for suggestions that, as Thurow puts it in his uncorrected reply, b (or P) was "judiciously selected" to maximize the error of his identification.[18]

Alternatively, for Thurow to insist that his belated introduction of the absurd relation $W=k^*+\gamma$ for zero savers confirms his model as a legitimate alternative throws doubt on the seriousness of his purpose. But the excessive indirectness of Motley and Morley's evidence may have helped fortify the impression that they were introducing an "alternative model" to Thurow's, when in fact they were providing a necessary complement to Thurow's initial structure to make the identification of $g(A)$ possible.

D

The tilting term $r(A)$ in the equation of income profiles was neglected in the previous sections, to accord with Motley and Morley's assumptions. With its reintroduction, equations (19) and (23) become:

$$(19')\quad \hat{y}_A = f(A)W - r(A)$$

$$(23')\quad \mathring{y}_A = [f(A)W - r(A)]\psi_A(Y_A)$$

The reasoning and assumptions of Section IIID lead to the modified expression:

$$(28')\quad \mathring{y}^*_A = G_4(A) = g(A)\,\frac{(1-p)f(A)+pr(A)}{f(A)-pg(A)}$$

The estimate $\hat{g}_A$ at each age of observation is derived from (28′) as

$$(29')\quad \hat{g}_A = \frac{\mathring{y}^*_A[F(A)+r(A)]}{p[y^*_A+r(A)]+(1-p)[F(A)+r(A)]}$$

where $F(A)=f(A)-r(A)$ is the mean income profile. Identification of $g(A)$ thus requires estimates of both p and $r(A)$ in addition to the mean income profile $F(A)$.

With the mean income at age M respecified as $F(M)$, the reasoning of Section IIIE leads to the estimating equation for p:

$$\mathring{c}_M/j(M) = (1-p) + p\,\frac{\mathring{y}_M + r(M)}{F(M)+r(M)}$$

i.e., observed incomes entering the ratios must be corrected by the factor $r(M)$. Since, however, the tilting factor is small in the middle range of the observed life span, the correction can be neglected and the value of p obtained in A can be retained.

---

[16] If, in addition to $H(W)=W$, we also set $\psi(W)=1$, the modified system (22)–(24) reduces to $f(A)/g(A)=\mathring{y}_A/\hat{c}_A=1$, and only determines ordinates of the two points of intersection of mean-income and mean-consumption profiles. The "shrinkage" of the observed consumption distribution in relation to that of income can certainly be explained in terms other than the statistical bias favored by proponents of the permanent-income hypothesis, i.e., in terms of a parameter e of $H(W)$ significantly different from one, as suggested in Section IIIc. In any case, strong assumptions concerning $H(W)$ are unnecessary for the resolution of the problem at hand.

[17] See Motley and Morley, pp. 740 and 742. Our parameters b and P have the same definition as in their article. Given (19), $h(W)=W$ in (18), and $e=1$, expression (27) reduces to

$$\hat{y}_A = \left(1-\frac{1}{b}\right)f(A) + \frac{1}{b}\mathring{y}_A$$

or

$$\hat{y}_A = (1-P)f(A) + P\mathring{y}_A$$

which is the equivalent of Motley's and Morley's equation (3).

[18] It evidently occurred to Thurow that Motley and Morley's equation (7), to which he refers in the uncorrected reply, can be read as $\mathring{\hat{y}}^* = (P-1)\overline{C}_t/(\gamma_t P-1)$ where $\overline{C}_t$ is our $g(A)$ and γ_t is $g(A)/f(A)$. The equality of $\mathring{\hat{y}}^*$ and $\overline{C}_t$ can only occur for $P=0$, and the bias of $\mathring{\hat{y}}^*$ as an estimate of $\overline{C}_t$ increases with the size of P. However, $P=0$ carries the absurd implication that a family's expected income and consumption at any given age are independent of its lifetime earnings, i.e., are equal, respectively, to average income and consumption.

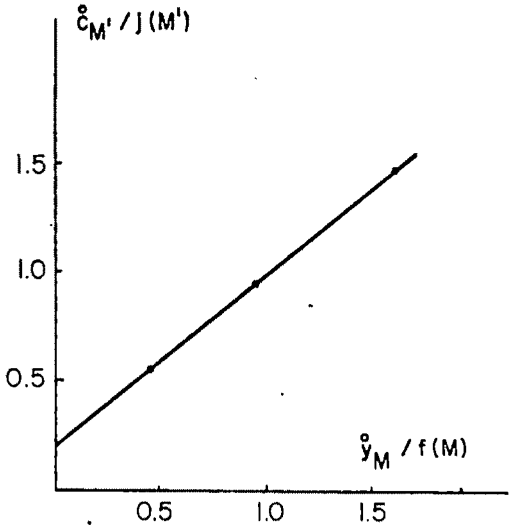


FIGURE 4

$$\overset{\circ}{c}_M/j(M) = 0.20 + 0.80\overset{\circ}{y}_M/f(M)$$

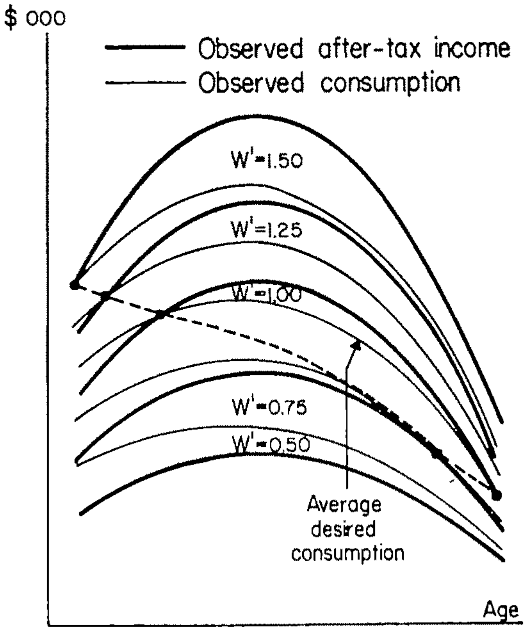


FIGURE 5. SIMULATION OF THUROW'S ESTIMATE

Introducing $p=0.80$ and *BLS* measures of $f(A)$ in relation (29), we obtain the estimate of $g(A)$ already calculated through an equivalent formula by Motley and Morley, p. 742, Figure 3. Comparison with the average profile of actual consumption suggests that discrepancies between desired and actual consumption are extremely minor and do not justify demands for drastic reform of *U.S.* lending policies as they affect younger age groups.

B

While the estimate of p presented in Section IVA may be questioned, it certainly reassures one that p is nowhere near zero. The fact of a considerable bias in Thurow's estimate of the desired-consumption profile is thus beyond question. For a visual illustration of its occurrence, a set of observations $\overset{\circ}{y}_A^*$ is graphed in Figure 5 under hypothetical specifications of the model. Starting with dome-shaped functions $f(A)$ and $g(A)$ describing $\overset{\circ}{y}_A$ and d_A as functions of A for $W=1$ and assuming $j(A)=g(A)$, we obtain four additional sets of observed income and actual consumption profiles for values of $\overset{\circ}{y}_A/f(A)=W'=0.50$, 0.75, 1.25, and 1.50. The $\overset{\circ}{y}$ profiles are obtained through direct multiplication of $f(A)$ ordinates by W'; the $\overset{\circ}{c}$ profiles are obtained through multiplication of $g(A)$—or $j(A)$—ordinates by $(0.20+0.80\ W')$. The function $\overset{\circ}{y}_v^*=G(A)$ goes through the intersections of each associated $\overset{\circ}{y}$ and $\overset{\circ}{c}$ curve (large dots) and is traced as a thick dotted line through interpolation. It obviously bears little relation to the shape of function $g(A)$.

C

Thurow's easy (if unwarranted) dismissal of similar findings by Motley and Morley can be traced to essentially one point, i.e., their initial assumption that $h(W)=W$ and $\theta(W)=1$, i.e., that the average propensity to consume is constant at each age and (much less significantly) that the biasing term $\theta(W)$ can be neglected. With $H(W)$ thus reduced to W (i.e., $e=1$), the parameter $p=e/b$ in

---

sumption. The assumption, therefore, is that $\overset{\circ}{c}_{M'}/j(M') = d_M/g(M)$ holds approximately true for sets (M, M') in the medium range of the observed life span. This (unnecessary) complication was excluded from the presentation in the text to achieve greater clarity.

$$\mathring{c}_A = g(A)H(W) \tag{25}$$
$$\text{with } H(W) = 1 - e + eW$$

$$\mathring{y}_A = f(A)\phi(W) \tag{26}$$
$$\text{with } \phi(W) = 1 - b + bW; \quad b > 1$$

Reducing (25) and (26) through elimination of W leads to

$$\mathring{c}_A/g(A) = (1 - p) + p\mathring{y}_A/f(A); \qquad p = e/b \tag{27}$$

and, in view of (24),

$$\mathring{y}^*_A = G_1(A) = g(A)\frac{(1 - p)f(A)}{f(A) - pg(A)} \tag{28}$$

This expression is of the same form as that obtained in Section IIIc with $r(A)=0$ and supports similar conclusions, i.e., the identification of the plot of $\mathring{y}_A$ with $g(A)$ is misleading unless $p=0$, or $e=0$ and $H(W)=1$. This, of course, no longer implies invariance of the desired-consumption profile with lifetime income; with $\theta(W)$ a variable function of W, $H(W)=h(W)\theta(W)=1$ implies that $h(W)$ also varies with W. Since, however, $\theta(W)$ is an increasing function of W, we would have to accept that $h(W)$ decreases as W rises, i.e., that the profile of desired consumption shifts downward as lifetime earnings increase. A correction of the plot to $\mathring{y}^*_A/k$ leads to identification problems similar to those uncovered in Section IIIc.

IV. Alternative Estimates of $g(A)$

A

It is apparent, then, that we must give up on Thurow's estimate. This, however, does not mean that $g(A)$ is beyond identification under proper use of the available information. Accepting the specifications in Section IIID, we can obtain an estimate, $\tilde{g}_A$, of $g(A)$ at each age of observation from (28):

$$\tilde{g}_A = \frac{\mathring{y}^*_A f(A)}{p\mathring{y}^*_A + (1 - p)f(A)} \tag{29}$$

All we need is an estimate of parameter p and a measure of average income, $f(A)$, at each age.

The reader is again referred to Figure 3 in which the *BLS* data available to Thurow were used to plot income and average consumption at the 10th, 50th, and 90th percentiles of the income distribution, for available age intervals. As a first step toward estimating p, write expressions (18) and (19) for the desired consumption and expected income of age M households with observed income and consumptions $\mathring{y}_M$, $\mathring{c}_M$:

$$d_M = g(M)h(W) \quad \text{and} \quad \hat{y}_M = f(M)W$$

and suppose that we expect $\hat{c}_M/j(M) = d_M/g(M)$, where $j(A)$ is average actual consumption as a function of age. The condition states that at age M intended consumption relative to its average is equal to desired consumption relative to its average.

The system above can then be written

$$\hat{c}_M = j(M)h(W) \quad \text{and} \quad \hat{y}_M = f(M)W$$

and the observed magnitudes, in view of (16) and (17), satisfy:

$$\mathring{c}_M = j(M)H(W) \quad \text{and} \quad \mathring{y}_M = f(M)\phi(W)$$

In correspondence with expression (27), we obtain

$$\mathring{c}_M/j(M) = (1 - p) + p\mathring{y}_M/f(M) \tag{30}$$

allowing p to be estimated by reference to several sets of observations of $\mathring{c}_M/j(M)$ and $\mathring{y}_M/f(M)$.

We assume that $\hat{c}_M/j(M) = d_M/f(M)$ holds approximately true in the medium range of the observed life span, and we arbitrarily set M at the age of maximum average income. Three sets of magnitudes ($\mathring{c}_M$, $\mathring{y}_M$) are thus determined by reference to the three sets of profiles in Figure 3, and the three corresponding sets $[\mathring{c}_M/j(M), \mathring{y}_M/f(M)]$ are obtained after reading average income and average consumption at age M from *BLS* tabulations (also the source of identification of M).

The three sets of ratios are plotted in Figure 4. The linear fit is almost perfect in the relevant range,[15] and we find:

---

[15] The ratios shown are based on measures of consumption at the age, M', of maximum average con-

D

A more careful reading of the empirical context does suggest, however, that conditions (2) or (11) as written are not relevant. As noted by Motley and Morley, Thurow does not have access to individual observations of zero-saving families. Rather, he obtains from Census sources large-sample averages of actual consumption expenditures and after-tax income for successive income intervals at different ages. It is well known from the "permanent income" literature that in the year of observation, households of age A with observed income and consumption $\overset{\circ}{y}_A$ and $\overset{\circ}{c}_A$ expect on the average an income $\hat{y}_A$, and intend a consumption $\hat{c}_A$, which differ from $\overset{\circ}{y}_A$ and $\overset{\circ}{c}_A$. The differences reflect the presence of "transitory" components of income and they are respectively related to the ratios $Y_A=\hat{y}_A/f(A)$ and $C_A=\hat{c}_A/j(A)$, where $f(A)$ and $j(A)$ are the income and consumption of an average family at age A.

Specifically:

$$(16) \qquad \hat{c}_A = \overset{\circ}{c}_A/\theta_A(C_A)$$

$$(17) \qquad \hat{y}_A = \overset{\circ}{y}_A/\psi_A(Y_A)$$

the functions $\psi_A(Y_A)$ and $\theta_A(C_A)$ increasing respectively with Y_A and C_A and taking value 1 when the variable is one.

Going back to the argument offered in Section II, it is apparent that statements such that "$d-c\geq 0$ under dissaving," or "actual dissaving cannot exceed desired dissaving," will not hold if *actual* income and consumption refer to observed rather than expected or intended quantities: If a household dissaves more than it intends because of a high negative transitory component of income, it is possible for observed dissaving to exceed the "desired" level. It follows that the proper zero-saving points to select for the purpose of identifying desired consumption are those for which $\hat{y}_A=\hat{c}_A$, where we expect $d_A=\hat{c}_A$. This, however, is not feasible, and we must follow Thurow and others in selecting observations for which $\overset{\circ}{y}_A=\overset{\circ}{c}_A$, or $\hat{y}_A=\hat{c}_A\theta_A(C_A)/\psi_A(Y_A)$, i.e., where intended consumption is different from expected income and, thus, intended saving is different from zero. Specifically, there is intended dissaving where zero saving is observed in the high range of income profiles and intended positive savings on low profiles.[13] We shall nevertheless assume $d_A=\hat{c}_A$ at points of zero observed savings ($\overset{\circ}{y}_A=\overset{\circ}{c}_A$), but as a provisional measure, and the resulting bias of desired-consumption estimates will be discussed in Section IVE.

In determining the possible effect of reinterpreted observations on the validation of Thurow's estimate, there is advantage in simplifying the expression of y_A (or $\hat{y}_A$) by elimination of the "tilting" term $r(A)$. This does not affect the substance of the conclusions, and it brings out strong similarities between this analysis and that of Motley and Morley. Under the new specification, and accepting $d_A=\hat{c}_A$ where zero savings are observed, we are dealing with a simultaneous system that includes (16) and (17) above and:

$$(18) \qquad d_A = g(A)h(W)$$

$$(19) \qquad \hat{y}_A = f(A)W$$

$$(20) \qquad \overset{\circ}{y}{}^*_A = \overset{\circ}{y}_A = \overset{\circ}{c}_A$$

$$(21) \qquad \hat{c}_A = d_A$$

where $\overset{\circ}{y}{}^*_A$ is the zero-saving observation at age A.

This reduces to:

$$(22) \qquad \overset{\circ}{c}_A = g(A)h(W)\theta_A(C_A)$$

$$(23) \qquad \overset{\circ}{y}_A = f(A)W\psi_A(Y_A)$$

$$(24) \qquad \overset{\circ}{y}{}^*_A = \overset{\circ}{y}_A = \overset{\circ}{c}_A$$

Under linear approximation of the terms $h(W)\theta_A(C_A)$ and $W\psi_A(Y_A)$, equations (22) and (23) become:[14]

---

[13] Because of the lesser sensitivity of consumption to transitory components of income, $\theta(C_A)$ increases with C_A much slower than $\psi(Y_A)$ does with Y_A. Writing the approximation $Y_A=C_A=Q_A$, the ratio $\theta_A(Q_A)/\psi_A(Q_A)$ is less than 1 for $Q_A>1$ and more than 1 for $Q_A<1$.

[14] Given the small expected range of $\theta_A(C_A)$, we assume it replaceable by a function $\theta(W)$ common to all ages and introduce the linear approximation $h(A)\theta_A(C_A)=h(A)\theta(W)=1-e+eW$.

We assume the function $\psi_A(Y_A)$ to be common to all ages and well approximated by the simple expression $\psi(Y_A)=1-b/Y_A+b$, $(b>1)$. Since, by (19), $Y_A=\hat{Y}_A/f(A)=W$, we have $W\psi_A(Y_A)=1-b+bW$.

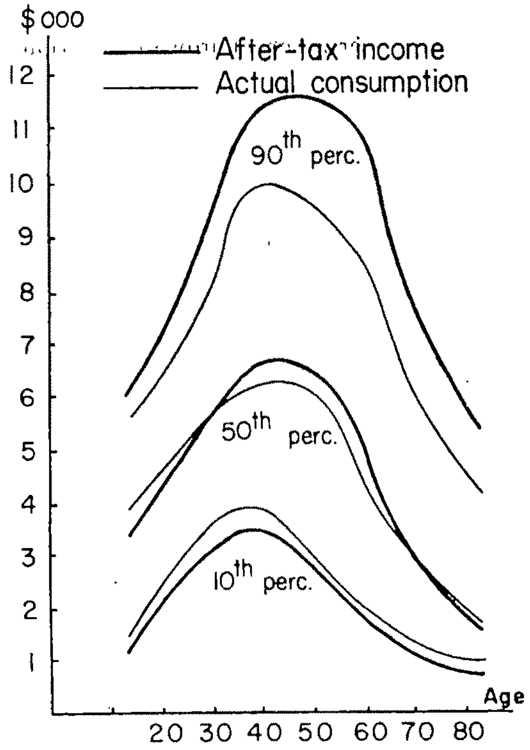


FIGURE 3. *BLS* DATA

Another is that intergenerational savings increase with lifetime earnings.[10]

Accordingly, we respecify the system developed in Section IIIB as follows:

$$d_A = g(A)h(W) + \eta \tag{9}$$

where η is independently distributed with mean zero, and the scaling factor $h(W)$ now appears as a function of W taking value 1 for $W=1$;

$$y_A = f(A)W - r(A) + \epsilon \tag{10}$$

where ϵ is independently distributed with mean zero, which simply reproduces expression (6).

Continuing to treat Thurow's measures as large-sample averages of individual observations, y_A^*, we retain (2) and (3) as:

$$\overset{*}{y}_A = y_A = c_A \tag{11}$$

$$d_A = c_A \tag{12}$$

Assuming $h(W)$ to be of the simple form[11]

$$h(W) = (1 - a) + aW \tag{13}$$

reduction of the system by elimination of W leads to:

$$\overset{*}{y}_A = G_2(A) + \zeta_2 \tag{14}$$

where

$$G_2(A) = g(A)\,\frac{(1 - a)f(A) + ar(A)}{f(A) - ag(A)} \tag{15}$$

$$\zeta_2 = \frac{\eta f(A) - \epsilon a g(A)}{f(A) - ag(A)}$$

Since ζ_2 is distributed with mean zero at each age, the plotting of a large-sample mean of y_A^* observations at successive ages will provide a correct estimate of $G_2(A)$. Again, however, the latter expression bears little resemblance to the function $g(A)$ sought by Thurow. It will only be equal to it if $a=0$; i.e., if $h(W)=1$. A correct identification of $g(A)$ with the plot would thus require that desired consumption be totally independent of lifetime income.

If, in more precise agreement with Thurow, the identification is with $\bar{y}_A^*/k$, we require:

$$g(A) = \frac{1}{k}\left[f(A)\,\frac{k - 1 + a}{a} - r(A)\right]$$

While we may accept an a priori specification of $g(A)$ in this form (with k treated as an unknown), we find that the equality of y_A and d_A at zero-saving points implies $y_A = f(A)h(W) + \epsilon'$, in which case zero-saving observations identify vertical lines at the intersection of $g(A)$ and $f(A)$ rather than the desired function $g(A)$.[12]

---

[10] Another possible bias is linked to intergenerational transfers: Low-income earners may receive more than their share of unreported parental gifts and free parental services.

[11] Equations (9) and (13) are consistent with explicit assumptions in Thurow's initial article, and with his "Erratum" statement to the extent that a is allowed to take value 1 and thus regenerate the system of Section IIIB.

[12] Replacing $g(A)$ by the proposed expression in (9), and writing $d_A = y_A$ (error terms excluded), reduction leads to $r(A) = f(A)(a-1)/a$.

ifications, we write:

$$y_A = f(A)W - r(A) + \epsilon \tag{6}$$

where ϵ is independently distributed with mean zero, which adequately describes the range of conceivable relations between income profile and wealth. The term $r(A)$ allows a "tilting" of the expected income profiles as a function of W, as suggested by observation (see Section IVD).

Joining equation (6) with (1), (2), and (3) specified in Section IIIA, i.e., replacing (4) by (6) in the original simultaneous system, reduction of the system by elimination of W leads to:

$$y^*_A = G_1(A) + \zeta_1 \tag{7}$$

where

$$G_1(A) = \frac{g(A)r(A)}{f(A) - g(A)} \tag{8}$$

$$\zeta_1 = \frac{\eta f(A) - \epsilon g(A)}{f(A) - g(A)}$$

Since ζ_1 is distributed randomly with mean zero at each age, the plotting of a large-sample mean of y^*_A observations at successive ages will provide a correct estimate of $G_1(A)$—and a homothetic adjustment of the plot by some factor $1/k$ will provide an estimate of $G_1(A)/k$. Clearly, the latter expression need not bear any resemblance to the function $g(A)$ sought by Thurow. In order to get $G_1(A)/k = g(A)$, we need:

$$g(A)r(A) = kg(A)[f(A) - g(A)]$$

or

$$g(A) = f(A) - r(A)/k$$

No relation of this type is assumed by Thurow, although the specification of $g(A)$ as a tilted version of $f(A)$, i.e., $g(A) = f(A) - r(A)/k$, k unknown, would not be totally unreasonable. However, the procedure is doomed by the fact that the equality of income and consumption at the observation points would either require a priori specification of W as equal to estimated k for all zero savers (which leaves us exactly where Thurow is) or impose elimination of the term $r(A)$ (in which case $g(A)$ is set identical to $f(A)$).[7] It is apparent that the discovery of a structure under which Thurow's estimation procedure would be proper requires the expenditure of some more subtlety.

C

In a further attempt at justifying Thurow, the reader is referred to Figure 3 in which the *BLS* data available to Thurow were used to plot average income and average actual consumption at the 10th, 50th, and 90th percentiles of the income distribution, for available age intervals.[8] It is apparent that the spread of consumption profiles is narrower than that of income profiles, i.e., that the overall level of consumption rises less than proportionately with the overall level of income.[9] While much of this "shrinkage" may represent a measurement bias attributable to averaging effects (see Section IIID), one possible explanation is that households with high lifetime earnings keep underestimating future income, and those with low lifetime earnings tend to overestimate future income.

---

[7] For zero savers, the equality of nonstochastic elements of y_A and d_A, as specified, implies:

$$f(A)W - r(A) = W[f(A) - r(A)/k]$$

$$r(A)\left(\frac{W}{k} - 1\right) = 0$$

$$r(A) = 0 \quad \text{or} \quad W = k$$

If we follow Thurow's practice of measuring lifetime incomes as simple sums of annual incomes over life, we must have $\sum r(A) = 0$ over the observed life span. With this specification and $W = k$ for zero savers, the system determines the unknown parameter k as Thurow's k^*.

[8] I am indebted to William Spaeth and Robert Gough for estimating these profiles from Thurow's sources. Spaeth's exploration of the data suggests that the identification of zero-saving income in each age group is far from unambiguous, and that the measures of after-tax income and consumption expenditure may suffer from serious biases, variable with age and income. For instance, the account-balancing difference in the younger age groups averages 5 percent of reported after-tax income.

[9] The overall excess saving revealed by the figure results primarily from the adoption of one of Thurow's two definitions of consumption, i.e., one that excludes insurance premiums and charitable contributions and, thus, formally incorporates the latter into savings. This definition is adhered to up to section IVE of the paper, when appropriate corrections are made.

come (zero actual saving). If we call y_A^* the income of a zero-saving family at age A, the observational constraint leads to the simultaneous condition:

$$\overset{*}{y_A} = y_A = c_A \tag{2}$$

where y_A is income and c_A actual consumption of the family.

Two additional conditions are introduced simultaneously with (1) and (2):

(a) For zero savers, desired consumption is equal to actual consumption:

$$d_A = c_A \tag{3}$$

(b) The lifetime income of zero savers is randomly distributed around some constant, i.e., the ratio of their lifetime income to average lifetime income is given by:

$$W = k^* + \gamma \tag{4}$$

where γ is independently distributed with mean zero. The constant k^* is defined and calculated as the ratio of summed averages of y_A^* observations in each age interval to summed averages of incomes in each age interval.

Reduction of the system (1), (2), (3), and (4) by elimination of W leads to:

$$\frac{\overset{*}{y_A}}{k^*} = g(A)(1 + \gamma') + \eta' \tag{5}$$

where $\gamma' = \gamma/k^*$ and $\eta' = \eta/k^*$ are independently distributed with mean zero.

If the above specification is correct, the plotting of $\bar{y}_A^*/k^*$ where $\bar{y}_A^*$ is a large-sample mean of observations y_A^* in age interval A should provide a correct estimate of $g(A)$. This indeed is the way in which Thurow obtains his estimate of $g(A)$, although, as indicated in Section IIID, the average income of zero savers which he measures at each age differs from $\bar{y}_A^*$.

Of the two conditions holding under zero saving, only the first, (3), is fully verbalized and justified by Thurow. The second, (4), is eventually made explicit in the "Erratum" contribution (Thurow's equations (1) and (2)), but never justified. This is sensible enough given the difficulty of the undertaking:[5] The notion that families found in a zero-saving situation at any age are all hovering around a level of lifetime income peculiar to them is, indeed, difficult to swallow. It is not inconceivable that some special form of the relation between income profiles, consumption profiles, and lifetime income could result in zero savers enjoying a common (expected) level of wealth. However, Thurow never describes with less than considerable ambiguity what income profiles occur at different levels of life-wealth,[6] and he does not, in any case, seem overly concerned by the oddity of condition (4).

While the hope of justifying condition (4) is too thin to pursue, we may, at least, attempt a check of whether any reasonable specification of the profiles could generate an estimation of $g(A)$ as $\bar{y}_A/k$, where k is a constant that may or may not equal k^*. Such an attempt is made—without success—in the next three sections, drawing on weak empirical information available from the *BLS* data.

B

As a first step in exploring promising spec-

---

[5] Motley and Morley make this point with great clarity, concluding that Thurow's procedure ". . . is valid only under the unstated (and almost certainly false) assumption that the lifetime income of the *zero-savers* is the same in every age group" (p. 739). In his original response, Thurow (1970) faults Motley and Morley for missing the assumption in his article. He calls attention to his assumption (2), and points out, p. 745, that he identified the bias likely to result from its dubious status. Assumption (2), however, merely states ". . . that individuals in older age brackets are similar in all relevant aspects with those in younger age brackets" (p. 326), and it refers to the fact that the average growth of earnings over time implies different lifetime earnings for individuals of different ages found to have the average income for their age in the year of observation. The assumption that such differences do not exist has, obviously, nothing to do with the assumption relating to zero savers, and Thurow's reference to it misses the point altogether.

[6] An unwary reading of Thurow's first article does yield a near-explicit assumption concerning income profiles and life-wealth; namely, that income profiles are derivable for one another by homothetic transformation in proportion to lifetime income (W). The context suggests, however, that the statement may refer instead to desired-consumption profiles in accordance with (1) above.

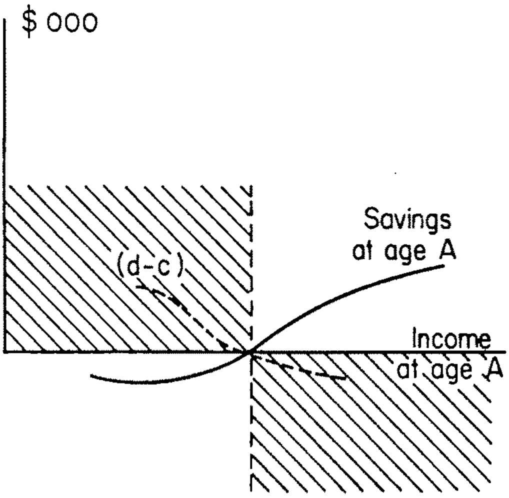


FIGURE 1

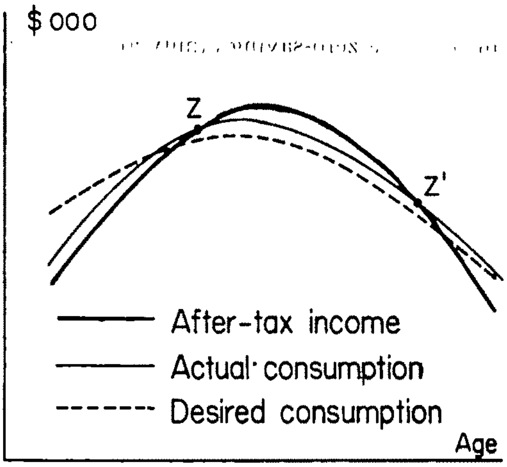


FIGURE 2. ILLUSTRATIVE PROFILES

age-profile of income, dissaving is at first rationed then becomes freely adjustable, actual consumption must rise with income until it reaches desired consumption, then continue up *above* the desired-consumption profile in view of the "excess" net assets created by former restrictions on dissaving. As illustrated in Figure 2, the intersection of actual consumption with income (zero saving) will generally occur above the desired consumption profile. The discrepancy (the excess of c over d at zero-saving point Z) depends on the extent of foregone dissaving in the earlier period and on the manner in which the household distributes additional consumption over the remainder of its life. Similarly, the zero-saving point Z' occurring later in life as income falls back may represent a consumption above the desired level if credit restrictions were effective in earlier years.

Short of developing more ambitious models, the selection of zero-savings observations can only be defended on the grounds that it guarantees a consistently positive bias of the desired-consumption measure, that this bias is likely to be smaller at the relatively low consumptions achieved at Z and Z' than in the peak period in between, and that, a posteriori, estimated desired consumption (when proper estimating methods are used) exceeds actual consumption by only small amounts and relatively late in the life cycle, thus assuring us that the biasing effect of foregone dissavings is weak over most of the profile.

III. A Case of Misidentification

The main contention of this review is that Thurow's analysis incorporates a fundamental error of identification, with the result that the estimated pattern of desired consumption represents nothing of the sort.

A

Although the precise nature of Thurow's assumptions remains clouded, he is quite specific in his last "Erratum" contribution concerning the relationship of desired consumption to lifetime income; i.e.,

$$d_A = g(A)W + \eta \tag{1}$$

where η is independently distributed with mean zero, d_A is desired annual consumption of the family at age A, $g(A)$ gives d_A as a function of age for a family of average lifetime income, and W is the family's lifetime income in relation to the average. Not only, then, are desired-consumption profiles at different levels of lifetime income obtainable from one another through homothetic transformation, but the proportionality factor is W itself.

Observations are constrained to situations in which actual consumption is equal to in-head are found by moving down the array of incomes at each age interval and stopping when saving (after-tax income *minus* consumption expenditure) falls to zero. The consumption expenditure (or after-tax income) thus identified is taken as the desired consumption of the average family when the age of its head is at the midpoint of the age interval. This procedure allows the estimation of as many points of the relationship as there are age intervals, and the remainder of the curve is obtained through interpolation. Thurow's plot of desired consumption is reproduced in Figure 6, where it can be compared with the actual-consumption profile.

The above procedure is justified by a seemingly obvious proposition: While people may be prevented from dissaving, there is no institutional constraint on saving, so that the after-tax income at which households in any age class show zero savings must represent the "desired" consumption expenditure of the average household at that age. A second look forces one to conclude that the statement can only be true of zero-saving households in the age class, not necessarily the "average" household. Further thinking suggests that by the reasoning offered, the observed consumption of households in a positive-saving situation should also be equal to its desired level. Finally it is not altogether clear that zero saving does imply equality of observed and desired consumption. Thurow recognizes these and other difficulties with his approach. One of them leads him to modify the estimate described in the previous paragraph: As the lifetime income of zero savers turns out to differ from the average, the curve is scaled along the consumption dimension so that aggregate desired consumption over life is equal to aggregate after-tax income. Overall, however, Thurow leans toward the position that the various problems he identifies do not seriously weaken his main conclusions, and their further elaboration by reviewers of his article leaves him undisturbed.

II. A Justification for the Zero-Saving Sample

Since both Thurow's analysis and the modifications proposed further down are based on the selection of zero-saving observations, some evaluation of the limits of validity of such a sample is in order.

If we designate actual consumption by c and desired consumption by d, we can state in the first approximation:

$d - c \geq 0$ if observed savings are negative

$d - c \leq 0$ if observed savings are positive

The first statement follows from the fact that nothing can force an individual to dissave more than he judges appropriate in the light of prospective incomes and interest rates, but that institutional borrowing restrictions may force him to dissave less (consume less) than what he perceives as optimum. The second statement follows from the observation that if desired dissaving has been thwarted in the first stage of life because of borrowing restrictions, the net asset position of the household has been increasing over its desired level (i.e., over the level achieved when dissaving proceeds optimally by reference to income and interest prospects), and the level of consumption selected in the later "saving" stage of the life cycle is higher (at any age, no less) than the level chosen under ever-perfect credit conditions. In other words, actual consumption may exceed desired consumption when observed savings are positive.

For successively higher household incomes at some given age, the rising curve in Figure 1 illustrates the change in savings. The shaded areas describe the field of possibilities for $(d-c)$ in accordance with the statements above. If in addition we assume that $(d-c)$ varies continuously with household income at the given age, the $(d-c)$ curve must cross the ordinate at the zero-saving point, i.e., zero-saving households are the only ones for which the equality of d and c is guaranteed. This presumably is Thurow's justification for his choice of observations, although his own formulation is excessively cryptic.

Unfortunately, the statements relating the sign of $(d-c)$ to that of savings do not hold strictly, as both Thurow and his critics have recognized. Assuming that in the typical

The Optimum Lifetime Distribution of Consumption Expenditures: Comment

By ANDRE DANIERE*

Two reviewers have already had a hand at exposing grievous errors in Lester Thurow's article concerned with the optimum consumption profile over life.[1] As will be recalled, Thurow attempted to estimate the pattern of consumption over life which the "average" family would select in the absence of institutional contraints on dissaving. This is the consumption pattern it would choose when faced with a given lifetime income and a rate of interest at which it may borrow or lend. Thurow finds such a pattern of consumption (which he calls *desired* consumption) to be substantially at variance with both the pattern of income and that of actual consumption over time, desired consumption (in the sense given) far exceeding actual consumption in early years. Thurow (1969) concludes:

> Based on the results of this analysis, lifetime welfare levels might be substantially increased if the constraints on lifetime income redistribution [mainly, constraints on long-term consumption borrowing] could be lifted. Consequently social planners should investigate methods of eliminating the institutional constraints. [p. 329]

Careful readers of the critical review offered by Brian Motley and Samuel Morley will be left in no doubt that the estimate of the desired-consumption profile offered by Thurow is wide of the mark—or, at the very least, that it bears no necessary relation to the curve he sought. There is evidence, however, that the too comprehensive and yet incomplete attack mounted by the reviewers, together with its airy dismissal by Thurow,[2] have left the average reader convinced of the essential validity of Thurow's conclusions. His results have already found their way into at least two recent reports concerned with loans to college students;[3] other sweeping proposals are likely to draw support from the new findings, possibly going all the way to subsidization of early-consumption loans (to make up for presumed attitudinal or informational handicaps of the young) at the cost of more worthy investment alternatives. There is thus some value in belaboring the point that Thurow's findings are misleading, this time concentrating on the least debatable and most damaging of his errors and calling it by its name; i.e., a classical case of misidentification. As a by-product, alternative estimates of the "desired-consumption" profile can be derived, taking advantage of Thurow's insights and refining the work of earlier reviewers.

I. Outline of Thurow's Estimation

The method of estimate used by Thurow is attractively simple and can therefore be described in simple form. His data base consists of survey figures showing average money income after taxes and average current consumption expenditure of families for different age-of-head and money-income intervals.[4] Points of the curve relating desired consumption of the average family to age of its is useless in any case, since the estimation of option price is no more difficult than the estimation of expected consumer's surplus. This assertion can be answered by a counterexample that indicates the sort of applied analysis I had in mind.

---

* Associate research professor, Boston College, and research economist, University of California, Berkeley. I am grateful for the help of William Spaeth and Robert Gough in processing data, and for that of two reviewers in suggesting corrections.

[1] The reviews, one by Kan Hua Young and the other by Brian Motley and Samuel Morley, appeared fifteen months after publication of Thurow's article in June 1969. All references in the present article are to the review of Motley and Morley.

[2] Thurow's first reply was followed, six months later, by an erratum correcting his earlier misinterpretation of the review by Motley and Morley, but reaffirming confidence in his own findings.

[3] See the works of Robert Hartman, p. 66, and D. Bruce Johnstone, p. 88, published in 1971 and 1972, respectively.

[4] See U.S. Bureau of Labor Statistics (*BLS*), Tables 14a–14g.

If one is concerned with the benefits attainable by preserving a wilderness area for recreational uses, a logical procedure if income effects are small is to estimate the aggregate demand curve for the area and to compute an aggregate surplus measure from it, a procedure followed by, for instance, John Krutilla et al. Since there are inevitably uncertainties attached to the level and shape of future demand functions, this approach yields, at best, aggregate expected consumer's surplus. I was concerned with the question of whether such a benefit measure should be scaled up because some of the associated uncertainty derives from the uncertain future preferences of potential demanders. My analysis indicates that, indeed, aggregate expected surplus is not a perfect measure of aggregate option price, the quantity one would most like to know. But, since I further showed that the bias (aggregate option value) can be of either sign, there is no general case for scaling up expected surplus estimates because preferences are uncertain. My suggestion amounted to the recommendation that such estimates be used without adjustment for this source of risk. Nothing in Bohm's comment has persuaded me to withdraw it or to make any other substantive alteration in my paper.[6]

---

[6] I would, however, strike the assertion that my definition of risk aversion is the "usual" one. Bohm's comment and a perusal of various texts and standard references have convinced me that no such definition exists.

REFERENCES

K. J. Arrow, *Essays in the Theory of Risk-Bearing,* Chicago 1971.

——— and A. C. Fisher, "Environmental Preservation, Uncertainty, and Irreversibility," *Quart. J. Econ.,* May 1974, *88,* 312–19.

P. Bohm, "Option Demand and Consumer's Surplus: Comment," *Amer. Econ. Rev.,* Sept. 1975, *65,* 733–36.

A. C. Fisher, J. V. Krutilla, and C. J. Cicchetti, "The Economics of Environmental Preservation: A Theoretical and Empirical Analysis," *Amer. Econ. Rev.,* Sept. 1972, *62,* 605–19.

M. Friedman and L. J. Savage, "The Utility Analysis of Choices Involving Risk," *J. Polit. Econ.,* Aug. 1948, *56,* 279–304.

C. Henry, "Investment Decisions under Uncertainty: The 'Irreversibility Effect'," *Amer. Econ. Rev.,* Dec. 1974, *64,* 1006–12.

C. Hildreth, "Expected Utility of Uncertain Ventures," *J. Amer. Statist. Assn.,* Mar. 1974, *69,* 9–17.

J. V. Krutilla et al., "Observations on the Economics of Irreplaceable Assets," in A. V. Kneese and B. T. Bower, eds., *Environmental Quality Analysis: Theory and Method in the Social Sciences,* Baltimore 1972.

R. D. Luce and D. H. Krantz, "Conditional Expected Utility," *Econometrica,* Mar. 1971, *39,* 253–72.

R. Schmalensee, "Option Demand and Consumer's Surplus: Valuing Price Changes Under Uncertainty," *Amer. Econ. Rev.,* Dec. 1972, *62,* 813–24.

possible states of nature, the U^i are the conditional indirect utility functions, the Y_i are the conditional incomes, and P is the vector of state-independent prices. Given (1), Bohm would prefer to say that the consumer is risk averse if and only if the U^i are all strictly concave in income, so that his definition relates directly to attitudes or preferences.[2] My Theorem 1 shows that a consumer who is risk averse in Bohm's sense will generally accept some fair gambles; this behavior can be ruled out only by also requiring the conditional marginal utilities of income, the U^i_y, to be equal at the point considered.

Thus, for example, even if the U^i are identical and strictly concave in income, the consumer will refuse all fair gambles if and only if the Y_i are also equal. Whether or not a consumer will agree to accept a particular gamble depends both on the structure of his preferences and on the sorts of risks he is bearing initially.[3] This point is easily made in the state preference framework, and, given the incompleteness of real-world contingent commodity markets, it is important in practice.

As I have indicated above and as Bohm's demonstration makes clear, my results do not depend on how risk aversion is defined, so that there is nothing to be gained in this context by arguing that one or the other of our two definitions captures the "true" meaning of the term. Since the word "aversion" has behavioral connotations, though, it might be clearer to use it in connection with my behavioral definition. Some other term, such as "concave preferences," would seem more appropriate to Bohm's attitude-oriented definition.[4]

One further aspect of Bohm's Section I requires a reply. Bohm seems to attach some normative significance to the income/price points considered in my Theorems 2–6, even going so far as to argue that some of them are "inappropriate." This is a confusion. In the absence of a well-specified set of trading opportunities in contingent commodities (such as that analyzed in Section III of my paper, to which Bohm apparently has no objection), there is no reason to expect the U^i_y to stand in any particular relation at any particular point, as Bohm in fact emphasizes. My results simply state that if, for whatever reason, the U^i_y are equal at the indicated points, the sign of option value can be determined. The points considered *were chosen* simply because they have this property, not because there is any welfare-theoretic reason to attach importance to them. Whether the conditional marginal utilities are or are not equal at any one of these points is purely a positive question, a question of fact.

II

Let me now consider Bohm's unhappiness with my suggestion that expected consumer's surplus be used as the best available approximation to option price. He asserts that there is something illegitimate about assuming that preferences can be represented by (1), so that my concept of expected consumer's surplus is not well founded. Formally, however, this assertion has no basis; see R. Duncan Luce and David H. Krantz for a detailed analysis of representations like (1), involving nonidentical conditional utility functions, and of the strong restrictions necessary to force the U^i to be identical.[5] Their work establishes that the state preference approach is indeed applicable to the case of uncertain future preferences.

Second, Bohm asserts that my suggestion

---

[2] Kenneth Arrow (p. 90) defines a risk averter as one *who will refuse* a fair gamble whenever his conditional incomes are equal. When the U_i are identical, this is equivalent to Bohm's definition, but in general it differs from both definitions discussed in the text.

[3] For an interesting analysis that explicitly considers both these elements, see Clifford Hildreth.

[4] I cannot resist making a final point about terminology. The terms "option value" and "option demand" would seem at first glance to have something to do with the value of keeping one's options open, of not reducing the set of possible or plausible actions. Hence, it would seem more natural to employ these terms in connection with option-foreclosing irreversible decisions, such as those considered in the extremely interesting analyses of Arrow and Anthony Fisher, Claude Henry, and Fisher et al.

[5] I am much more sympathetic to the argument that as a matter of fact, most consumer behavior is not sufficiently consistent and clear-headed to be adequately described by expected utility maximization, but that argument is not directly relevant here.

Option Demand and Consumer's Surplus: Reply

By RICHARD SCHMALENSEE*

In his comment, Peter Bohm does not take issue with my principal result, that the sign of option value is generally ambiguous. Rather, he is unhappy with my derivation of this result. In particular, he objects to the way I used the term "risk aversion." His second main objection is to my suggestion that expected values of consumers' surpluses be used as the best available approximation to option prices. I shall answer these in turn.

I

It should be noted at the outset that neither my results nor the proofs used to establish them are called into question by Bohm's comment, and Bohm neither extends nor generalizes my findings. He objects only to my interpretation of these results in terms of risk aversion. Thus the statement at the close of his Section I that option value "may thus be positive, negative or zero, as Schmalensee argued, but not quite for the reasons he chose to give" is totally misleading.

Bohm apparently misunderstands the role of the definition of risk aversion in my analysis. This may be because, as is often the case, the order in which results were presented in my paper did not correspond to the order in which I obtained them.[1] From the definition of option price and the assumed concavity of the conditional indirect utility functions, I was first led to inequalities like his equations (6) and (7) and my (12). As my main goal was to show that the sign of option value is ambiguous, it was necessary to exhibit plausible conditions under which option price exceeded or fell short of expected surplus. Given the relevant inequalities, this is easy, and the conditions simply involve equality of the conditional marginal utilities of income at the income/price points considered in my Theorems 2–6. Since, as Bohm agrees, the conditional marginal utilities may stand in any relation to each other at any point under the assumptions of my analysis, it follows that the sign of option value is ambiguous. Thus, none of the implications of my Section II would have been changed had I replaced "is risk averse" everywhere by "has equal conditional marginal utilities of income." Had I made this change, dropping Theorem 1 and the attendant discussion, I doubt that Bohm would have felt compelled to offer his comment.

In fact, the paper assumed roughly this form in its first draft. I found the presentation unsatisfactory, however, without some indication of the observable consequences, and thus the economic meaning, of having equal conditional marginal utilities of income. Theorem 1 and the association of this equality with risk aversion arose from the search for such consequences. This theorem provides an economic interpretation of the hypotheses of my Theorems 2–6, and it also points up the important distinction between attitudes toward risk and behavior under risk.

I considered a consumer to be risk averse at some income/price point *if and only if he* would refuse all fair gambles at that point, so that my definition relates directly to behavior. This is precisely the definition (of "preference for certainty") given in Section 3 of the pioneering study of Milton Friedman and Leonard J. Savage. Suppose now that the consumer's preferences can be represented by the indirect utility function

$$(1) \qquad V(Y_1, \ldots, Y_N; P) = \sum_{i=1}^{N} \Pi_i U^i(Y_i, P)$$

where the Π_i are the probabilities of the N

---

* University of California, San Diego.

[1] I realize that very little interest, if any, attaches to the manner in which I wrote the paper in question. I discuss it only because Bohm seems to infer, probably from the structure of that paper, that my main interest was in the effects of risk aversion, and this is simply not true.

II

Given that option price may deviate from the expected consumer's surplus in one direction or the other, it may still be argued—as Schmalensee does—that the expected consumer's surplus is the best or even the only approximation to the option price that is available. Now, to calculate the expected consumer's surplus, the decision maker needs to know in principle both the set of π_i for each consumer and the individual consumer's surplus in each state. The decision maker can hardly be expected to be clairvoyant enough to know the π_i, and there is no market behavior to study that reveals these probabilities. In addition, the SC_i is something the decision maker might not be able to find out about. Estimating the consumer's surplus entails the same well-known problems as measuring other access values such as the *WTP* for (access to) public goods—or *OP*, for that matter. Thus, from a practical viewpoint estimating *OP* does not add any new dimensions to the measurement problems that always exist for cases where an estimate of the consumer's surplus is required.

We just argued that it may be difficult for the government to estimate the π_i for each consumer. This is actually a superfluous argument, as we shall now question the relevance of Schmalensee's assumption that the π_i may be known to anyone. It is, as we have seen, different preference functions that constitute the different states. Thus, one state may be characterized by, say, a sudden whim to travel to Yellowstone National Park next summer—to take Schmalensee's illustration. The remarkable thing about this consumer who does not know what he wants to do next summer is that he is capable of knowing and specifying (a) the complete set of his alternative preferences next summer, and (b) the (approximate) probabilities of these alternatives.

Rejecting this assumption as being unrealistic, it is no longer possible to determine the expected consumer's surplus as defined here ($\Sigma\pi_i SC_i$). The option price is therefore the *only* measure of the benefit side of the investment that can conceivably be determined—by sales of access rights, by interviews, by government "introspection," or other imperfect approaches.

III

To sum up, I can agree with Schmalensee's results, given his state-preference approach, on the following points: (a) option values may be either positive or negative in the presence of uncertain future preferences and risk aversion; (b) there may be nonzero option values also in the presence of risk neutrality. However, Schmalensee reached these conclusions on the basis of a definition of risk aversion that I cannot accept. In particular, I have argued that the relevant definition of risk aversion does not require that the U_y^i be identical in all states.

But, more important, I have argued that Schmalensee's state-preference approach is not applicable to the case of uncertain future preferences. As to Schmalensee's remark that " . . . the expected value of the consumers' surpluses ought to be employed as the best available approximation to the sum of their option prices . . . " (p. 823), I have pointed out (a) that it overlooks the fact that the estimation problems are essentially the same for option prices and consumers' surplus, and (b) that there is no information available to the decision maker about how to calculate the *expected* value of the consumers' surplus. Schmalensee's concluding remark that " . . . there would appear to be no practical way to obtain superior estimates" (p. 823) simply has no support in his analysis, as he has not dealt with the practical problems of how to estimate option prices, consumer's surplus, and similar kinds of access values, given that there is no market solution to the problem.

REFERENCES

C. J. Cicchetti and A. M. Freeman III, "Option Demand and Consumer Surplus: Further Comment," *Quart. J. Econ.*, Aug. 1971, *85*, 528–39.

R. Schmalensee, "Option Demand and Consumer's Surplus: Valuing Price Changes Under Uncertainty," *Amer. Econ. Rev.*, Dec. 1972, *62*, 813–24.

willing to pay for access to the commodity in question at a given price, ***given which state, i.e., preference function, holds.*** But the U_y^i certainly does affect OP, as OP is the maximum price the consumer is willing to pay "today" in order to have access to the commodity at the given price ***in all conceivable "future" states.*** And in the latter case it is naturally relevant to take into account how income changes are valued in the various states. Leaving the U_y^i out would be as if an exporter selling at a given future price in foreign currency did not care about the fact that exchange rates differ in different states.

(b) *Strictly concave U^i functions (risk aversion):*

$$OP > SC_{OP}^{w}$$

If the U^i instead are strictly concave functions, the option price will be larger than the right-hand side in (6) and smaller than the right-hand side in (7). The reason why (6) gives an underestimate and (7) an overestimate of OP is clear from looking at (3′)–(5). Equation (3′) as it stands determines OP; the SC_i are given by (1), and all other variables (except OP) are exogenous. Then, introducing the inequalities (4) and (5) into (3′), we obviously get a lower limit and an upper limit to OP, respectively.

Having said this we may drop (7) out of the picture. The interesting expression is (6) where U_y^i is taken at the point from which the consumer makes his evaluation. Remember that we are asking him how much he can abstain from today in terms of an OP and enter the future state, whatever it may be, with a disposable income of $Y-OP$, without being worse off. He will be at $Y-SC_i$ only if he does not pay an option price—and that is another story. We do not ask him about the maximum amount he is willing to pay provided he does *not* pay that amount.

(c) $OP \gtreqless \Sigma\pi_i SC_i$

What is the relationship between SC_{OP}^{w} in (6), the weighted expected consumer's surplus, and the expected consumer's surplus, $\Sigma\pi_i SC_i$? We know that they are equal when $U_y^i = U_y^j$ for all i and j. The same is true for another special case, viz., $SC_i = SC_j$ for all i and j. In such cases, OP equals or exceeds $\Sigma\pi_i SC_i$ depending on whether the U^i are linear or strictly concave as shown by Schmalensee. In general, however, this relationship does not hold, as we have already pointed out. We may give an example where $\Sigma\pi_i SC_i$ is smaller than SC_{OP}^{w}, and hence, where $OP > \Sigma\pi_i SC_i$ regardless of the alternative shapes of the U^i functions. Take the two-state case where $\pi^1 = \pi^2 = 1/2$ and where $SC_2 = 0$, i.e., where the commodity in question would not be demanded in state 2. (This is the case analyzed by Cicchetti and Freeman.) Thus, for

$$SC_{OP}^{w} = \frac{\Sigma\pi_i SC_i U_y^i}{\Sigma\pi_i U_y^i} = \frac{1/2\, SC_1 U_y^1}{1/2\,(U_y^1 + U_y^2)}$$

$$> \Sigma\pi_i SC_i = 1/2\, SC_1$$

we have that $\quad U_y^1 > 1/2\,(U_y^1 + U_y^2)$

or that $\quad U_y^1 > U_y^2$

It is quite possible of course that this requirement is fulfilled. The same is true for the opposite, i.e., $U_y^1 < U_y^2$, which would mean that $OP \geq SC_{OP}^{w} < \Sigma\pi_i SC_i$. In a case such as this, OP may be smaller than $\Sigma\pi_i SC_i$ (for example, when the U^i functions are linear).

Thus, we have that $OP \gtreqless \Sigma\pi_i SC_i$. If the difference between the two sides is defined as an *option value*, it may thus be positive, negative, or zero, as Schmalensee argued, but not quite for the reasons he chose to give. Schmalensee actually argues that the sign of the option value in the presence of risk aversion depends on whether the U_y^i are measured at $Y-OP$ or at $Y-SC_i$; as we have already pointed out the latter point of measurement is not relevant. (Given Schmalensee's definition of risk aversion and given the consumer's surplus defined as the compensating variation, the sign is actually unambiguously determined as nonnegative.) As we have shown here, the sign of the option value is indeterminate and depends on the relationship between the U_y^i in the different states.

$$(3') \quad \Sigma\pi_i[U^i(Y - OP, P^*) - U^i(Y - SC_i, P^*)] = 0$$

Let us assume that U^i is a concave function so that, in particular,

$$(4) \quad U^i(Y - OP, P^*) - U^i(Y - SC_i, P^*) \geq (SC_i - OP)U^i_y(Y - OP, P^*)$$

$$(5) \quad U^i(Y - OP, P^*) - U^i(Y - SC_i, P^*) \leq (SC_i - OP)U^i_y(Y - SC_i, P^*)$$

Equation (4) inserted into (3′) gives

$$\Sigma\pi_i(SC_i - OP)U^i_y(Y - OP, P^*) \leq 0$$

or

$$(6) \; OP \geq \frac{\Sigma\pi_i SC_i U^i_y(Y - OP, P^*)}{\Sigma\pi_i U^i_y(Y - OP, P^*)} (\equiv SC^w_{OP})$$

and (5) inserted into (3′) gives

$$(7) \; OP \leq \frac{\Sigma\pi_i SC_i U^i_y(Y - SC_i, P^*)}{\Sigma\pi_i U^i_y(Y - SC_i, P^*)} (\equiv SC^w_{SC_i})$$

(SC^w_{OP} and $SC^w_{SC_i}$ stand for the weighted expected consumer's surplus at $Y-OP$ and $Y-SC_i$, respectively). Schmalensee concentrates on the cases where the U^i_y in (6) *or* U^i_y in (7) are the same for all states. (Cicchetti-Freeman make a similar assumption.) If so, the right-hand side becomes $\Sigma\pi_i SC_i$. Hence, OP is either $\geq$ or $\leq \Sigma\pi_i SC_i$, depending on at which point the U^i_y are equal for all i.

Now, it is difficult to see that this type of equality has any particular interest other than as a possible special case. (In spite of the emphasis in Schmalensee's analysis, this seems to be his opinion as well.) What constitutes different states in this problem are different preference functions; hence, as there are no bounds on "how much" they may differ, it may be considered natural that they differ also with respect to U^i_y at any particular point such as $(Y-OP, P^*)$.

The reason why Schmalensee deals almost exclusively with the special case just mentioned is that he wants to find out about the effects of risk aversion and argues that risk aversion exists at a given point, for example, $(Y-OP, P^*)$, if and only if the U^i_y are the same at that point. His definition of risk aversion may be questioned, however. Schmalensee uses the "usual definition of risk aversion" (p. 815) saying that all accepted gambles with payoff w_i in state i have the property $\Sigma\pi_i w_i > 0$. But this definition stems from the case where there is a unique preference function. If there are several, as is the case here, individual behavior must be analyzed in utility space, not in money space; in other words, the marginal utility of income in each state plays a role in determining behavior.

Let us restate the argument briefly as follows: In contrast to the case where only (Y^i, P), $Y^i = Y + w_i$, differ among states we now have different $u^i(Y^i, P)$ among states. Hence, we have a utility function $U[u^i(Y^i, P)]$ instead of $U(Y^i, P)$ and expected utility as $\Sigma\pi_i U[u^i(Y^i, P)]$ instead of $\Sigma\pi_i U(Y^i, P)$. While risk aversion in the latter case is defined by $\Sigma\pi_i w_i > 0$, the relevant definition would now seem to be that $\Sigma\pi_i w_i u^i_y > 0$. If so, risk aversion is present in both cases if U is strictly concave. If, moreover, u^i is concave (which is natural to assume given the assumption of risk aversion) we have that $U^i(Y^i, P) = U[u^i(Y^i, P)]$ is strictly concave. Hence it follows from an assumption of risk aversion that U^i is strictly concave. Likewise, we may add, risk neutrality is given by U and the u^i being linear, which implies that the U^i are linear.

Given now that u^i_y and hence, that $U^i_y = U' u^i_y$ typically differ among states we can establish the following results:

(a) *Risk neutrality:*

$$OP = SC^w_{OP} \neq \Sigma\pi_i SC_i$$

If the U^i are linear functions, the right-hand sides of (6) and (7) will be the same and equal to OP. Thus, (6) or (7)—at equality—determines the option price in the absence of risk aversion. Note, however, that OP is not equal to the "expected consumer's surplus" $(\Sigma\pi_i SC_i)$ under risk neutrality. Of course, the U^i_y have no relevance for the future consumer's surpluses, the SC_i, as this is the maximum price the consumer would be

Option Demand and Consumer's Surplus: Comment

By PETER BOHM*

The concept of option values—briefly, the idea that consumers under uncertainty would be willing to pay more than the expected consumers' surplus for the option to buy a certain commodity in a future period—has for some time been a source of disagreement. Some economists have argued that option values are like the Emperor's new clothes—there just are no such things. Others argue that option values may exist and if so, that they are positive. Again, others argue that option values may equally well be negative.

An attempt to clarify this issue was recently made in an article by Richard Schmalensee. Although his contribution sheds light on important aspects of the option value problem, the relevance, and hence, the practical conclusions of his approach can be questioned. The implications of Schmalensee's results are summarized in his concluding two sentences, where he suggests that "... when tastes are the main source of uncertainty, the expected value of consumers' surpluses ought to be employed as the best available approximation to the sum of their option prices.... Benefits will sometimes be underestimated and sometimes overestimated with this procedure, but there would appear to be no practical way to obtain superior estimates" (p. 823). We shall argue that this piece of advice to investing government agencies is unacceptable. Our conclusions are presented in Section III. To begin with, we present a brief summary of Schmalensee's analysis.

I

Schmalensee intends to show that it is not generally true as argued by Charles Cicchetti and A. Myrick Freeman and others that risk aversion in the presence of uncertain future demand for a good leads to positive option values. He argues instead that option values may be positive or negative according to the specific nature of risk aversion.

The background may be described as follows. Given any investment project of the decreasing-cost type, we know that optimal prices will (may) not cover costs. In order to arrive at an investment decision we need to know how much consumers are willing to pay for (future) access to the project commodity. A concrete example may be a new line of public transportation that can be introduced next year (with fares on a given level). This can be regarded as a comparison between two sets of prices P and P^*, the latter differing from the former only to the extent that introduction of the project is equivalent to a "price reduction" of the project commodity. According to Schmalensee, the individual consumer is uncertain about his future preferences in the sense that he does not know which of a given set of alternative preference functions (future states) will hold, but he can attach given probabilities, π_i, to each state i. (As the possibility of future incomes being different in different states is irrelevant in the present context, we regard income as given.)

The consumer's surplus (SC_i) in each future state (alternative future preference function) is defined here as the compensated variation,

$$(1)\quad U^i(Y - SC_i, P^*) = U^i(Y, P)$$

The maximum willingness to pay (WTP) now, i.e., the option price (OP), is defined by

$$(2)\quad \Sigma\pi_i U^i(Y - OP, P^*) = \Sigma\pi_i U^i(Y, P)$$

Inserting (1) in (2), we get

$$(3)\quad \Sigma\pi_i U^i(Y - OP, P^*) = \Sigma\pi_i U^i(Y - SC_i, P^*)$$

Equation (3) can be rewritten in the following form:

---

* Department of economics, University of Stockholm. I wish to thank Claude Henry for his helpful remarks on an earlier version of this note.

rights should get all of the gains from trade, not just part.

III. Greenwood, Ingene, and Horsfield's Comment

The comment by Peter Greenwood, Charles Ingene, and James Horsfield (hereafter G-I-H) is in many ways similar to the G-S comment and to a degree suffers from similar specification problems. Like the G-S comment, G-I-H are specifying the bribes and payments in terms of changes in profit levels. Unlike G-S, the G-I-H comment attempts to deal explicitly with the bargaining situation. As with G-S, the bargaining is costless. G-I-H models develop supply and demand functions, the interaction of which is used as the determinant of the bargaining process. There are two interesting problems with this "solution." First, you have two firms in a bilateral bargaining situation where each firm is acting as a price taker. It would seem that either you have bilateral monopoly or you have firms acting as price takers, but not both simultaneously. Secondly, like the G-S paper, in the bribe and liability situations there seems to be a compromising of the property rights. The firm with the property rights has *all* the bargaining power, while the other ends up purchasing concessions. While there are of course conditions under which the property right can be "divided" and traded, the essential point (in the G-I-H comment) of *how* such division and transactions occur is based upon the inconsistency pointed out above.

IV. A Concluding Note

Our analysis was begun with the objective of looking at the Coase theorem in a specific, consistent neoclassical framework. G-S and G-I-H, in reaching conclusions contrary to ours, have also attempted to make use of a neoclassical framework. Their models, while interesting, are based on specifications and behavioral postulates which are either logically and internally inconsistent or not fully and properly developed. This latter criticism applies also to the Coelho paper. Thus, while we stand by the results contained in our original article, we anticipate continued research on the problems of externalities and the development of more fully specified and consistent bargaining models.

REFERENCES

W. J. Baumol, "On Taxation and the Control of Externalities," *Amer. Econ. Rev.*, June 1972, *62*, 307–22.

J. M. Buchanan and W. C. Stubblebine, "Externality," *Economica*, Nov. 1962, *29*, 371–84.

G. Calabresi, "Transaction Costs, Resource Allocation and Liability Rules," *J. Law Econ.*, Apr. 1968, *11*, 67–73.

R. Coase, "The Problem of Social Cost," *J. Law Econ.*, Oct. 1960, *3*, 1–44.

T. D. Crocker, "Conceptual Choice," *Natur. Resources J.*, Oct. 1973, *13*, 561–77.

G. Daly, "The Coase Theorem: Assumptions, Applications, and Ambiguities," *Econ. Inquiry*, 1974, *12*, 203–13.

R. C. d'Arge, "Introduction to the Coase Theorem Symposium," *Natur. Resources J.*, Oct. 1973, *13*, 557–60.

F. T. Dolbear, "On the Theory of Optimum Externality," *Amer. Econ. Rev.*, Mar. 1967, *57*, 90–103.

H. H. Liebhafsky, "The Problem of Social Cost—An Alternative Approach," *Natur. Resources J.*, Oct. 1973, *13*, 615–76.

G. W. Nutter, "The Coase Theorem on Social Cost: A Footnote," *J. Law Econ.*, Oct. 1968, *11*, 503–7.

R. Turvey, "On Divergences between Social Cost and Private Cost," *Economica*, Aug. 1963, *30*, 309–13.

graphical analysis to attempt to show that the allocation of resources is invariant to the form of the cost function. Unfortunately, Coelho's analysis is deficient in a number of respects, and while the analysis that he does present seems to support his conclusion, in fact, a corrected presentation would generate our original conclusions. Referring to Coelho's graphical models, we will indicate how Coelho's analysis is faulty.

Using Figure 1, showing the total revenue and total cost curves facing firm *A*, and Figure 3, showing the total revenue and total cost curves for firm *B*, Coelho derives a locus showing the profit possibilities (in Figure 4) of the two firms if merged. This analysis is correct in showing the potential profit tradeoffs between the two processes which are possible under joint ownership. However, as drawn, Coelho's graphs do not deal with the bargaining situations and leave unclear the entire adjustment process by which equilibrium at point *N* in Figure 4 is attained if the firms are not merged. In essence, Coelho is asserting his conclusions, not proving them. In a complete analysis, Figure 1 should indicate not only the revenue based on price times quantity for firm *A*, but also the potential revenue to be gained from extracting a bribe from firm *B*. Similarly, the curves for *B* in Figure 3 should indicate the effects of money transfers between firm *A* and firm *B*, as a function of the alternative output levels. Furthermore, and this is the crucial point, the relevant total revenue and total cost curves for *both A* and *B* will depend on whether the cost function is separable or nonseparable. Coelho is correct in showing (in Figure 2) that constant parametric shifts in the total cost curve leave unchanged the profit maximizing output of the firm whose total cost curve has shifted. However, whether the shift is constant for each level of output or varies with the output depends on the separability properties and can be seen only by looking at the marginal revenue and marginal cost curves (or alternatively, the first-order conditions). This difference is at the heart of our original analysis, and the interested reader can reconstruct Coelho's curves to take into account the effects of bribes and payments on the cost functions *and* the revenue functions in the separable and nonseparable cases.

II. Gifford and Stone's Comment

In contrast to Coelho, Adam Gifford and Courtenay Stone (hereafter G-S) have presented models which are alternatives to ours and allow concrete analysis and proof. Their models specify the bribes and damage payments in terms of foregone profits rather than imposed costs as in our models. In addition their models implicitly assume that the two parties bargain costlessly and that the actual payment exchanged is some convex combination of the maximum payment the payer will give up and the minimum payment the recipient will accept. Our objections are with this aspect of their specification. The two firms being modeled are cast in a situation of bilateral monopoly. One can simply not assume some resolution in a costless manner. A costless resolution of differences or splitting of the gains from trade goes against the entire existing theory of bilateral monopoly and simply buries the eminent question in bilateral monopoly of bargaining and the gains versus the costs.

Our second objection with the G-S analysis rests with the disposition of the gains from trade. As Coelho has stressed, there are gains from trade to be had from cooperation in a situation of external diseconomies between two firms. In the two situations under contention, one or the other of the parties has the property rights assigned to them, either explicitly as when the damager is made liable or implicitly when there is no rule otherwise. For both of these situations the G-S analysis results in a splitting of the gains from trade between the two parties. It would seem that this flies in the face of the property rights assignment. The party who has the property rights is in a position to capture all of the gains from trade simply by virtue of the initial placement of the property rights. The party that does not have the property rights has no bargaining power, i.e., he is initially indifferent between paying a bribe and bearing the externally imposed cost. Thus the party with the property

Externalities, Liability, Separability, and Resource Allocation: Reply

By JAMES R. MARCHAND AND KEITH P. RUSSELL*

Ronald Coase's paper, "The Problem of Social Cost" has become a classic article in the literature on externalities. Out of that paper has come the "Coase Theorem." Coase himself never stated a theorem as such in his original paper, but subsequent papers, texts, and discussions have developed statements of Coase's main ideas in terms of a theorem. (Coase himself may not have preferred .this; perhaps proposition is better than theorem.) Statements of the Coase Theorem or proposition vary slightly from source to source, but almost all cover the same concepts. Generally most economists, the authors and their critics included, could come to some agreement on a statement of a "loose Coase Theorem." This theorem would more than likely be inclusive of the following ideas about a two-firm diseconomy situation:

1) There are possible gains from trade.
2) If the firms can merge, overall profits to the merged firm will be larger than the sum of the profits of the two separate firms.
3) If the two firms can successfully costlessly bargain, transact, or exchange, Pareto optimal output is likely to attain.

However, though economists might agree on a general Coase Theorem since Coase's paper appeared, there have been many papers attempting to examine Coase's ideas in terms of specifics and neoclassical analysis.[1] These papers have dealt with the question of how the two parties bargain, exchange, or transact given various definitions of property rights or liability and scope for negotiation. Our original paper and the comments made upon it are a part of this continuing discourse. With this frame of reference, the remainder of our reply will speak to the comments in response to our original analysis, first in general, and then in terms of the specifics of each paper.

In general one can summarize the thrust of our critics' charges with the term "model misspecification." Each of the three comments makes the claim that certain of the M-R models are misspecified. In organizational terms each of the comments is very similar in structure. Each of the three papers present alternative models which the authors contend are properly specified. On the basis of these models each of the comments either asserts or derives conclusions which are opposite to ours and sustaining of the Coase Theorem. None of the authors finds fault with our mathematical analysis. Our critics' theme seems to be that models are misspecified when they do not yield the right conclusions.

Our main response to our critics is the same as their critique of us We feel that their models are misspecified. Like our critics, we feel that their modeling and analysis is correct mathematically speaking. Where we feel that the misspecification occurs is in terms of the assumptions that our critics make. Generally speaking, it appears that our critics are making assumptions which are not in the spirit of the original situation envisioned by Coase or are making assumptions which are internally contradictory. The rest of this reply is devoted to a discussion of how these general comments apply specifically to each comment.

I. Coelho's Comment

In his comment, Philip Coelho uses a

---

* Assistant professors of economics, California State University, Northridge. We are indebted to Jack Hirshleifer for helpful comments. Responsibility for errors remains with us.

[1] Early papers were by G. Calabresi, F. T. Dolbear, G. W. Nutter, and R. Turvey. More recent papers are by W. Baumol, T. Crocker, G. Daly, R. d'Arge, and H. H. Liebhafsky.

may be a direct function of B's output, thereby giving A cause to bribe B to reduce its output. A will so demand until the marginal cost of its bribe equals the marginal reduction in legal payments.

Deriving A's demand for reduction of q_2, where P_3 is the price per unit reduction paid by A and q_2^1 is B's output in M-R's case C, corresponds with solving the following problem:

$$\begin{aligned}\underset{q_1;q_2}{\text{maximize}}\ \pi_1 = P_1q_1 - A(q_1)& \\ - \left[B(q_1, q_2) - B(0, q_2)\right]& \\ - P_3(\overset{1}{q_2} - q_2)&\end{aligned}$$

subject to $q_1 \geq 0$; $q_2 \geq 0$; P_3

A's demand schedule is characterized by the following conditions:

$$P_1 = A_1(q_1) + B_1(q_1, q_2) \tag{6}$$

$$P_3 = B_2(q_1, q_2) - B_2(0, q_2) \tag{7}$$

Since when costs are separable $B_2(q_1, q_2) \equiv B_2(0, q_2)$, A has no demand for bribing B; no bribing occurs. The allocation resulting is given by M-R (equations (5′) and (6′)), and again this is the optimal allocation.

In either case, B is willing to supply reduction to the point where the marginal bribe received equals the marginal pre-bribery profit foregone. B's "supply" of reduction is found as the result of the following problem:

$$\begin{aligned}\underset{q_2}{\text{maximize}}\ \pi_2 = P_2q_2 - B(0, q_2)& \\ + P_3(\overset{1}{q_2} - q_2)&\end{aligned}$$

subject to $q_2 \geq 0$; P_3

B's "supply" is given by:

$$P_3 = P_2 - B_2(0, q_2) \tag{8}$$

Combining (6), (7), and (8) yields the optimal allocation. Thus, it has been shown that the allocation of resources is independent of both the legal framework and the nature of the cost functions when transactions are costless, and both are price takers.

Throughout this comment we have made a pair of assumptions, one explicit, the other implicit. The latter has been that all bribery results are marginal, that is, neither firm will discontinue operations as the result of bribery payments or legal sanctions. The former assumption has been that both firms act as price takers in the bribery market. If such a situation does not hold, if one or both of them attempt to exert market power, the optimal allocation of resources is in general destroyed.

We analyze one case of market power, leaving other cases to the interested reader. The legal system permits A to impose uncompensated external costs upon B, and A is aware of B's demand function. A's decision is summarized by:

$$\begin{aligned}\underset{q_1;q_2}{\text{maximize}}\ \pi_1 = P_1q_1 - A(q_1)& \\ + B_1(q_1, q_2)(\hat{q}_1 - q_1)& \\ + \lambda\left[P_2 - B_2(q_1, q_2)\right]&\end{aligned}$$

subject to $q_1 \geq 0$; $q_2 \geq 0$

The solution of this problem is the allocation of resources which is characterized by the following conditions:

$$P_1 = A_1 + B_1 - (\hat{q}_1 - q_1)(B_{11}B_{22} - B_{12}^2)/B_{22} \tag{9}$$

$$P_2 = B_2 \tag{10}$$

These conditions imply the optimal allocation only when the last term of (9) is zero. The optimal allocation clearly cannot be guaranteed.

However, this does not mean that the optimal allocation cannot be reached. As long as further gains can be made there is incentive for (costless) negotiations to continue, and there are always gains to be shared when the allocation is suboptimal.

We hope this comment on the topic of externalities will aid in the understanding of Coase's basic argument about the nature of markets.

REFERENCES

R. Coase, "The Problem of Social Cost," *J. Law Econ.*, Oct. 1960, *3*, 1–44.

J. Marchand and K. Russell, "Externalities, Liability, Separability, and Resource Allocation," *Amer. Econ. Rev.*, Sept. 1973, *63*, 611–620.

Externalities, Liability, Separability, and Resource Allocation: Comment

By Peter Greenwood, Charles Ingene, and James Horsfield*

In a recent article in this *Review*, James Marchand and Keith Russell (M-R) attempt to demonstrate that " . . . the neutrality of liability on resource allocation holds only in the restricted case of additively separable cost functions" (p. 611). M-R reach this conclusion because they incorrectly specify the form that a bribe between two firms—the injurer (A) and the injured (B)—will take. When the bribery market is correctly specified, the form of the cost functions has no effect on the allocation of resources; further, the allocation of resources does not depend upon the delimitation of liability by the legal system. In this comment we establish a neoclassical market framework within which we analyze the various situations described by M-R; this analysis supports what they term "a 'generalized' Coase theorem" (p. 618).

The nature of the trades is dependent on the legal framework. Specifically, assume that the legal system permits A to impose external costs on B without requiring compensation. B has cause to bribe A in order to obtain a reduction in A's output. B will so "demand" to the point where the marginal cost of the bribe is equated with its marginal (external) cost reduction. The derivation of B's demand corresponds with the solution of the following problem; where P_3 is the parametric bribe paid by B per unit of reduction of q_1, $\hat{q}_1$ is A's pre-bribery output (M-R case A), and the remaining notation is given by M-R:

$$\underset{q_1;q_2}{\text{maximize}}\ \pi_2 = P_2q_2 - B(q_1, q_2) - P_3(\hat{q}_1 - q_1)$$

$$\text{subject to } q_1 \geq 0; \quad q_2 \geq 0; \quad P_3$$

B's demand schedule is implied by the following conditions:

$$(1) \qquad P_2 = B_2(q_1, q_2)$$

$$(2) \qquad P_3 = B_1(q_1, q_2)$$

Firm A will "supply" B with a reduction in quantity until the marginal bribe received is equal to the marginal pre-bribery profit foregone. An alternative but equivalent statement is that A will produce until the price it receives for its product is just equal to the sum of its marginal cost and the marginal bribe. The derivation of A's supply schedule corresponds with the following problem:

$$\underset{q_1}{\text{maximize}}\ \pi_1 = P_1q_1 - A(q_1) + P_3(\hat{q}_1 - q_1)$$

$$\text{subject to } q_1 \geq 0; \quad P_3$$

The answer to this problem yields the following supply function:

$$(3) \qquad P_3 = P_1 - A_1(q_1)$$

Provided that both A and B approach the bribery market with a competitive spirit (as price takers), A will reduce its output until B's marginal (external) cost equals A's marginal pre-bribery profit foregone. This represents an optimal allocation of resources, as can be seen by combining equations (1), (2), and (3) to obtain (M-R's equations (7) and (8)):

$$(4) \qquad P_1 = A_1(\overset{*}{q}_1) + B_1(\overset{*}{q}_1, \overset{*}{q}_2)$$

$$(5) \qquad P_2 = B_2(\overset{*}{q}_1, \overset{*}{q}_2)$$

This is the efficient allocation; the result is invariant with respect to the nature of the cost functions.

We now consider a legal system which does not permit the imposition of uncompensated external costs. A's legally required payments sis in Section II indicates. The value of Marchand and Russell's analysis is that while clearly incorrect in their specification of the inapplicability of Coase's results for the nonseparable cost situation, their results serve as a reminder that legal definitions when different from economic ones, will affect resource allocation and economic efficiency. An additional caveat from their analysis is that resource efficiency is not aided by incorrectly specifying the economic damage imposed by negative externalities.

---

* Assistant professor of economics, Southeastern Massachusetts University; acting assistant professor of economics, University of Oklahoma; and staff economist, Economic Research Service, U.S.D.A., respectively.

REFERENCES

A. Gifford, Jr. and C. C. Stone, "Externalities, Liability and the Coase Theorem: A Mathematical Analysis," *Western Econ. J.*, Sept. 1973, *11*, 260–69.

J. R. Marchand and K. P. Russell, "Externalities, Liability, Separability, and Resource Allocation," *Amer. Econ. Rev.*, Sept. 1973, *63*, 611–20.

which is, of course, firm A's reduction in profits from its decrease in output. The maximum bribe that firm B will offer firm A for such a reduction is

$$C^B = p_2q_2 - B(q_1, q_2) - p_2\hat{q}_2 + B(\hat{q}_1, \hat{q}_2)$$

which is firm B's increase in profits consequent to firm A's reduction in output and firm B's expansion to its new optimal rate, q_2. C^A and C^B represent the limits to the actual bribe, C, which can be represented as

$$C = \alpha C^A + (1 - \alpha)C^B; \qquad \alpha \in (0, 1)$$

where the level of α is determined by the bargaining strategies of the two firms.[3] From the objective functions for the two firms:

Firm A: $\underset{\{q_1\}}{\text{Maximize:}}$ $p_1q_1 - A(q_1) + C$

subject to $q_1 \geq 0$

Firm B: $\underset{\{q_2\}}{\text{Maximize:}}$ $p_2q_2 - B(q_1, q_2) - C$

subject to $q_2 \geq 0$

come the following solutions:

$$p_1 = A_1(\overset{*}{q}_1) + B_1(\overset{*}{q}_1, \overset{*}{q}_2) \tag{12}$$

$$p_2 = B_2(\overset{*}{q}_1, \overset{*}{q}_2) \tag{13}$$

which are identical to M-R's socially optimal solutions (7) and (8).

Environment C. If firm A is "fully liable" for the imposition of the externality, then firm B has the property right to the profit resulting from the maximization problem:

Firm B: $\underset{\{q_2\}}{\text{Maximize:}}$ $p_2q_2 - B(0, q_2)$

subject to $q_2 \geq 0$

which yields firm B's maximum profit when

$$p_2 = B_2(0, \bar{q}_2) \tag{14}$$

where $\bar{q}_2$ is firm B's optimal output rate *in the absence of the externality*. Since firm B must be guaranteed at least this profit level, the minimum payment that firm B will require firm A to pay in order to produce some positive output level q_1 is

$$L^B = p_2\bar{q}_2 - B(0, \bar{q}_2) - p_2q_2 + B(q_1, q_2)$$

where, again, q_2 is firm B's optimal output when firm A produces q_1. The maximum liability that firm A will incur is the difference between its profits here and its next most profitable opportunity (different location or industry). As this alternative profit is an opportunity cost to firm A, it appears in firm A's cost function, making firm A's maximum bribe $L^A = p_1q_1 - A(q_1)$. The actual liability incurred[4] is

$$L = \tilde{\alpha}L^A + (1 - \tilde{\alpha})L^B; \qquad \tilde{\alpha} \in (0, 1)$$

The maximization problem for the firms is now

Firm A: $\underset{\{q_1\}}{\text{Maximize:}}$ $p_1q_1 - A(q_1) - L$

subject to $q_1 \geq 0$

Firm B: $\underset{\{q_2\}}{\text{Maximize:}}$ $p_2q_2 - B(q_1, q_2) + L$

subject to $q_2 \geq 0$

resulting in the following solutions:

$$p_1 = A_1(\overset{*}{q}_1) + B_1(\overset{*}{q}_1, \overset{*}{q}_2) \tag{15}$$

$$p_2 = B_2(\overset{*}{q}_1, \overset{*}{q}_2) \tag{16}$$

Again, the levels of output are identical with those obtained for the previous problem (12) and (13) and for M-R's social optimal output rates (7) and (8). Different *economic* liability rules do *not* result in different social resource allocation in the nonseparable cost situation.

III. Liability Definitions and Resource Use

That Coase's Theorem holds irrespective of either the property right assignment or the form of the cost function is not surprising. Profit-maximizing firms will respond to changes in their profit possibilities by adjusting the levels of output to maximize the total value of output when property rights are exchangeable at zero costs—as the analy-

---

[3] Ruling out a corner solution necessarily requires that $C^B > C^A$. The possibility of corner solutions, however, in no way affects Coase's conclusions. For a complete discussion, see the authors.

[4] Similar to the discussion in fn. 2, ruling out a corner solution requires that $L^A > L^B$.

$$+ [B(\hat{q}_1, q_2) - B(q_1, q_2)]$$

subject to $q_1 \geq 0$

Firm B: $$\underset{\{q_2\}}{\text{Maximize:}} \quad p_2q_2 - B(q_1, q_2) - [B(\hat{q}_1, q_2) - B(q_1, q_2)]$$

subject to $q_2 \geq 0$

resulting in outputs q_1'' and q_2'' from the solution values:

$$(3) \qquad p_1 = A_1(q_1'') + B_1(q_1'', q_2'')$$

$$(4) \qquad p_2 = B_2(\hat{q}_1, q_2'')$$

C. Imposition of Liability. (Firm A imposes an externality on firm B; firm B has the (property) right to have firm A compensate firm B for the damage imposed.) M-R state: "... a law or judge ... [declares] ... firm A liable for the extra costs experienced by firm B in a state of nature. Liability is therefore $L = B(q_1, q_2) - B(0, q_2)$" (p. 614). The relevant decisions for the firms are now

Firm A: $$\underset{\{q_1\}}{\text{Maximize:}} \quad p_1q_1 - A(q_1) - [B(q_1, q_2) - B(0, q_2)]$$

subject to $q_1 \geq 0$

Firm B: $$\underset{\{q_2\}}{\text{Maximize:}} \quad p_2q_2 - B(q_1, q_2) + [B(q_1, q_2) - B(0, q_2)]$$

subject to $q_2 \geq 0$

yielding the solution values

$$(5) \qquad p_1 = A_1(q_1') + B_1(q_1', q_2')$$

$$(6) \qquad p_2 = B_2(0, q_2')$$

D. Centralized Decision Making. (Firm A and firm B merge, thereby internalizing the externality.) The relevant objective function for the new super-firm is:

$$\underset{\{q_1, q_2\}}{\text{Maximize:}} \quad p_1q_1 + p_2q_2 - A(q_1) - B(q_1, q_2)$$

subject to $q_1 \geq 0$; $q_2 \geq 0$

which yields the socially optimal (for resource allocation) outputs

$$(7) \qquad p_1 = A_1(\overset{*}{q}_1) + B_1(\overset{*}{q}_1, \overset{*}{q}_2)$$

$$(8) \qquad p_2 = B_2(\overset{*}{q}_1, \overset{*}{q}_2)$$

M-R then show (for their assumptions about the second-order partial derivatives) that:

$$(9) \qquad \hat{q}_1 > \overset{*}{q}_1 \quad \text{and} \quad \hat{q}_2 < \overset{*}{q}_2$$

$$(10) \qquad q_1' < \overset{*}{q}_1 \quad \text{and} \quad q_2' > \overset{*}{q}_2$$

$$(11) \qquad q_1'' > \overset{*}{q}_1 \quad \text{and} \quad q_2'' < \overset{*}{q}_2$$

M-R conclude that "In the more complicated, yet more realistic world where production externalities involve nonseparable cost functions, the allocation of resources is not independent of the legal framework. As the inequalities (9)–(11) show, the allocation of resources depends on the definition of liability" (p. 618).

We now demonstrate that while the allocation of resources does indeed depend upon the *legal* definition of liability, Coase's Theorem applies generally, regardless of the nature of the cost function.

II. The Generalized Coase Theorem[2]

The Coase Theorem, as derived from an economic analysis of the impact of different property-right (liability) assignments on resource allocation, is clearly based on an *economic* definition of liability. Given the competitive situation under analysis, achieving the greatest value of production involves using resources in their highest-valued (i.e., most profitable) use. Therefore, the economic cost (damage) that firm A inflicts on firm B is the decrease in firm B's *profits* consequent to the external effect generated by firm A. It is easy to show that by using foregone profits as the cost or damage for liability purposes, Coase's Theorem holds regardless of the form of the cost function. Reconsider M-R's environments B and C, using the economic definition of liability.

Environment B. If firm A is assigned the right to generate the externality, it will reduce its output (and also the external effect) from $\hat{q}_1$ to q_1 for any bribe larger than

$$C^A = p_1\hat{q}_1 - A(\hat{q}_1) - p_1q_1 + A(q_1)$$

---

[2] This analysis is based on an extended presentation developed elsewhere by the authors.

Externalities, Liability, Separability, and Resource Allocation: Comment

By ADAM GIFFORD, JR. AND COURTENAY C. STONE[*]

In a recent article in this *Review*, James Marchand and Keith Russell deny the generality of Ronald Coase's classic analysis of the two-firm externality situation where transactions cost and income-redistribution effects on demands are zero. Specifically, Marchand and Russell (M-R) state that "... Coase's conclusion regarding the neutrality of liability on resource allocation holds only in the restricted case of additively separable cost functions" (p. 611). Their analysis indicates that in the case of nonseparable cost functions, not only will the levels of output differ depending upon the allocation of liability (thereby challenging Coase's symmetry results), but also that the levels of output will differ from the socially optimal output levels regardless of the liability assignment (challenging Coase's implications for efficient resource allocation).

We show that M-R's seemingly startling conclusions result not from the form of the cost function, but from their own peculiar specifications of the opportunity cost of the externality under different property-right assignments. We first summarize the M-R analysis for the nonseparable cost function case. This is followed by the corrected analysis which demonstrates that Coase's conclusions are independent of the form of the cost function. Finally, some aspects of the importance of different definitions of liability and property rights are examined.

I. M-R's Analysis

Marchand and Russell state:

> Consider two firms, A and B, each a member of a different competitive industry, whose productive processes interact. Denote A and B outputs as q_1 and q_2 and their total cost functions as $A(q_1)$ and $B(q_1, q_2)$, respectively. (These functions relate the costs of the inputs purchased directly by the firms to the levels of output indicated in the arguments.) Assume that $A_1>0$, $B_1>0$, $B_2>0$, $A(0)=0$, and $B(q_1, 0)=0$. [p. 612]

From this initial situation M-R develop four alternative scenarios, or "environments."

A. Simple Profit Maximization. (Firm A imposes an externality on firm B; firm B can't find firm A.) From the objective functions for the two firms:

Firm A: $\underset{\{q_1\}}{\text{Maximize:}}\quad p_1q_1 - A(q_1)$

subject to $q_1 \geq 0$

Firm B: $\underset{\{q_2\}}{\text{Maximize:}}\quad p_2q_2 - B(q_1, q_2)$

subject to $q_2 \geq 0$

the resulting first-order conditions for a maximum involve the solutions:[1]

$$(1)\qquad p_1 = A_1(\hat{q}_1)$$

$$(2)\qquad p_2 = B_2(\hat{q}_1, \hat{q}_2)$$

B. Expanded View of Profit Maximization. (Firm A imposes an externality on firm B and has the (property) right to do so; firm B offers a bribe (compensation) to firm A to reduce the externality.) M-R state: "... the compensation which B would be willing to pay ... is equal to the cost saving B receives as a result of A's producing less than $\hat{q}_1$; that is, compensation would be $C=(B(\hat{q}_1, q_2) - B(q_1, q_2))$" (p. 613). From the relevant objective functions (including the compensation):

Firm A: $\underset{\{q_1\}}{\text{Maximize:}}\quad p_1q_1 - A(q_1)$

---

[*] Assistant professor, Southern Methodist University, and associate professor, California State University, Northridge, respectively.

[1] The following equations and numbering up to and including equation (11) correspond to the M-R equations and numbering.

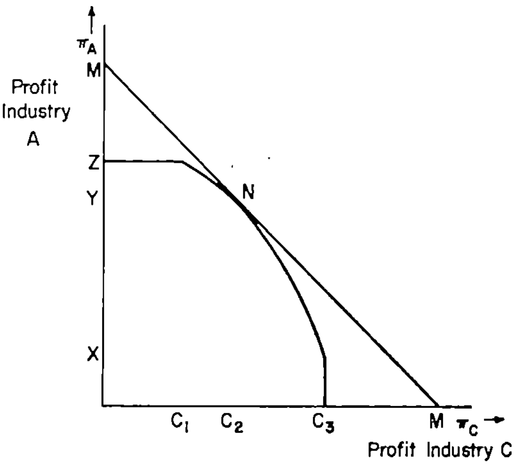


FIGURE 4

in keeping with the basic assumption behind the Coase Theorem—that negotiations between the two parties are feasible. Figure 4 illustrates the point. The curve ZNC_3 is the locus of points which show the maximum profit available to industry C for given levels of profit in industry A. This curve will be called the profit-opportunity locus. The line $(M-M)$ tangent to the profit-opportunity locus at N is the highest indifference curve the entrepreneur and society can reach. The indifference curve MM is a straight line with a slope of -1, indicating that neither society nor the entrepreneurs care from which industry the dollars of profit come. As long as point N is not reached, both parties can be made better off (i.e., each can have a little more profit) by negotiating (trading) until point N is reached.

The mathematics that Marchand and Russell present are flawless. Their economic logic leaves something to be desired. In Section IB, "An Expanded View of Profit Maximization" (p. 613), the entrepreneurs are presumed ignorant of the true nature of their joint production functions. What else can explain the failure of one of the businessmen to offer to buy out the other (for a dollar more than the capitalized value of his firm) and then proceed to maximize joint profits? As I have pointed out, merger is not the only solution; with reference to Figure 4 a system of side payments *can and will be made* assuming no transaction costs and knowledge of the profit-opportunity locus. If one party lacks the knowledge of the profit-opportunity locus and the other has it, then an outright purchase of the ignorant industry will be made by the knowledgeable industry. Only if both parties are ignorant will the Marchand and Russell conclusions hold.

The line of reasoning pursued here is not dependent upon the shape of the profit-opportunity locus. Any frontier of whatever shape will do equally well. It is dependent upon the shape of the indifference curve. Our optimum point N will only be reached if the indifference curve for profits is a straight line with a slope of -1. In essence, the Coase Theorem holds no matter what the nature of the cost function is, given 1) an absence of ignorance by at least one of the parties and 2) no transaction costs. It seems likely that if the respective industries are unaware of the nature of the externalities then all other parties would be equally ignorant. In this final case the question of optimum subsidies or taxes becomes moot.

Obviously, in the real world there are transaction costs, and knowledge does cost resources. Whether we should rely upon voluntary or governmental action in a specific case depends upon the equally obvious point: the difference between private and governmental costs of acquiring knowledge and of enacting transactions.

REFERENCES

R. Coase, "The Problem of Social Cost," *J. Law Econ.*, Oct. 1960, *3*, 1–44.

O. Davis and A. Whinston, "Externalities, Welfare and the Theory of Games," *J. Polit. Econ.*, June 1962, *70*, 241–62.

J. P. Marchand and K. P. Russell, "Externalities, Liability, Separability, and Resource Allocation," *Amer. Econ. Rev.*, Sept. 1973, *63*, 611–20.

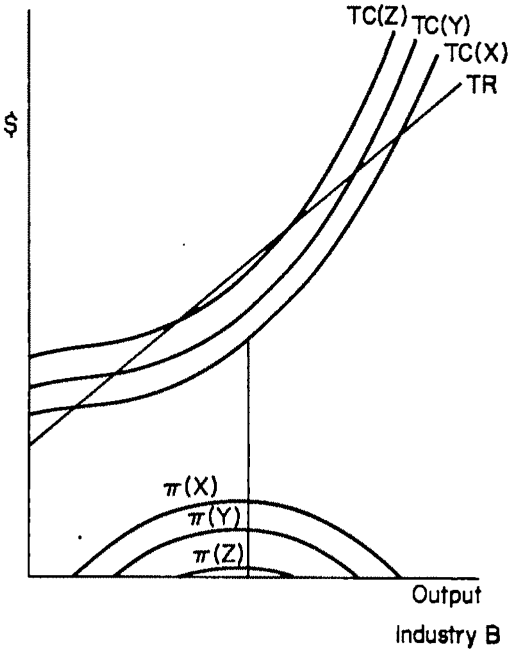


FIGURE 2

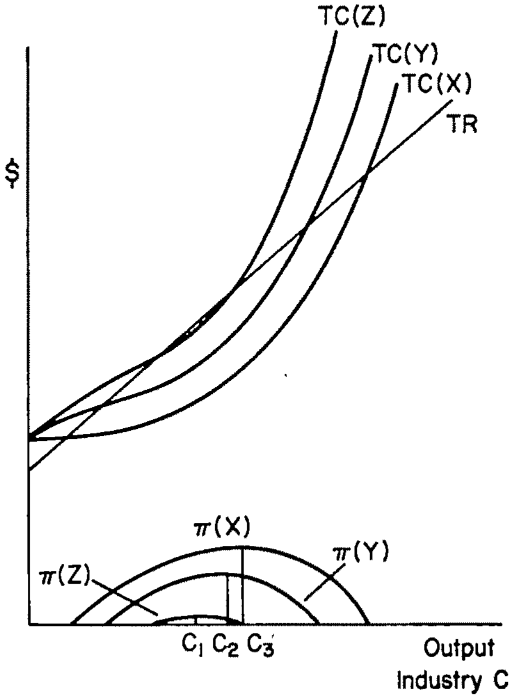


FIGURE 3

the additional costs were added to industry *B*.

Figure 3 depicts an industry (*C*) that has a nonseparable cost function. (In order to simplify the following discussion, no reference will be made to Figure 2 nor to industry *B*; we will deal with solely two industries, *A* and *C*. The addition of a third industry does not change our analysis, it adds a third dimension, however.) Industry *C*'s total cost function depends upon the level of output it produces and the level of output in industry *A*. The output of industry *A* enters multiplicatively in the cost function of *C*; i.e., the cost function is nonseparable.

This is demonstrated by observing that the profit-maximizing output of industry *C* changes as output in *A* changes. If industry *A*'s output is *Z* (the output that would maximize industry *A*'s profits in the absence of side payments to or from *C*), then the profit function for *C* will be $\pi(Z)$ and the profit-maximizing output will be C_1 in industry *C*. If industry *A*'s output is at *X* (the maximum output industry *A* can produce without generating externalities), industry *C* will produce output C_3 and achieve its maximum profit on the profit function $\pi(X)$.

Suppose that outputs *Y* of industry *A* and C_2 of the *C* industry are the outputs that maximize joint industry profits. Marchand and Russell argue that if industry *C* is forced to pay a bribe to prevent some damage, industry *C* will produce less than the joint profit-maximizing solution and industry *A* will produce more. In the case where liability for damage rests upon industry *A* their results are reversed. However, their argument explicitly constrains the entrepreneurs to these outcomes.[3] This is neither realistic nor

---

[3] They state: "Finally, *it is possible* that one might argue against our results on the grounds that we did not allow the individuals concerned sufficient scope to negotiate. That is, no matter what liability rule is in effect any arrangements are preliminary in nature and lead to more complex negotiation such that it is possible for individuals to merge and attain the jointly optimal solution" (pp. 617–18, emphasis added). This is precisely the point of Coase's article! Yet, their mathematics constrain the entrepreneurs to a nonoptimal point. The situation is analogous to Cournot's solution to the duopoly problem: the mathematics are impeccable, the economic logic motivating the individuals described by the math is ludicrous.

Externalities, Liability, Separability, and Resource Allocation: Comment

By PHILIP R. P. COELHO*

In a recent article in this *Review*, James Marchand and Keith Russell have disputed the generality of the Coase Theorem. The essence of this theorem is that in the absence of negotiation and transaction costs, an externality will have no effect upon the efficient allocation of resources regardless of where the liability for damage rests.

The argument of Marchand and Russell stresses the difference between separable and nonseparable cost functions. The Coase Theorem is shown by the authors to be true for separable cost functions but not for nonseparable functions.[1] I see no merit in the distinction. This note will show that given the Coase assumptions, the Coase Theorem is valid no matter what the nature of the cost function. In the following, a heuristic approach to the problem will be employed. It is my opinion that the strict mathematics is what led the authors into incorrect economic analysis.

A separable cost function is distinguished from a nonseparable one in that it can be divided into additive terms with only one variable per term. Consequently, when a separable cost function is differentiated with respect to its (own) output, marginal cost is solely a function of output. A nonseparable cost function cannot be segregated into additive terms with only one variable per term. Consequently, the marginal cost function of a nonseparable cost function will have other variables in it besides its (own) output.

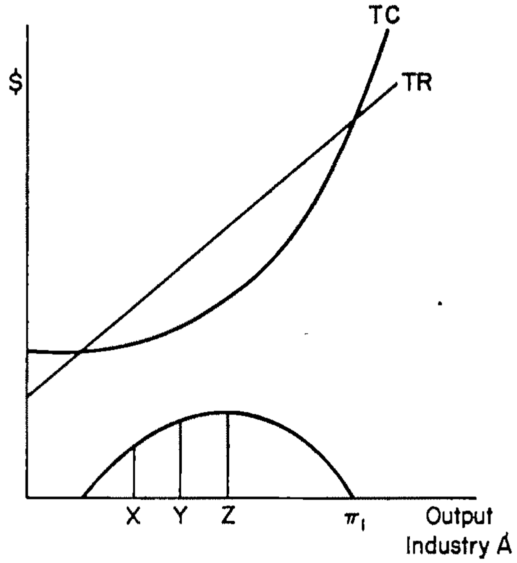


FIGURE 1

The following figures illustrate the distinction. Figure 1 depicts industry A which is an industry that is not subject to externalities and consequently has a separable cost function. TR and TC are abbreviations for total revenue and total cost, π is the total profit function. Figure 2 depicts an industry that has a separable cost function and that is subject to externalities.[2] As the output of industry A increases, costs in industry B increase. However, the profit-maximizing output does not change in industry B; the only change that occurs is that the cost curves shift up by a constant for each increase in cost imposed by the output of industry A. Figures 1 and 2 will be recognized as the "normal" total cost, total revenue, and total profit curves found in microeconomics texts. Figure 2 differs from Figure 1 by varying constants added to the total cost curve. Consequently, the marginal conditions for profit-maximizing output will not differ from what they were before at level $1/2\ P_m$ where P_m is the highest willingness-to-pay for any demanded consumption unit. The marginal cost curve intersects the marginal benefit curve at points associated with capacities Q_1 and Q_m. But capacity Q_1 minimizes net benefits because the marginal cost curve intersects the marginal benefit curve from above. Therefore, the correct intersection is the one associated with capacity Q_m since here the marginal cost curve intersects the marginal benefit curve from below. However, the area under the marginal cost curve from 0 to Q_m is just equal, by construction, to the area under the marginal benefit curve from 0 to Q_m, so at capacity Q_m net benefits are zero. This means we should just be indifferent between a capacity of zero or a capacity of Q_m. On the other hand, for any marginal cost curve that is constant at a level that is less than $1/2\ P_m$, the area under the marginal cost curve will be less than that enclosed by the marginal benefit curve, so net benefits are greater than zero, and the optimum capacity is Q_m. Similarly, for any marginal cost curve that is constant at a level that is higher than $1/2\ P_m$ the optimum capacity is zero.

---

* Washington State University.

[1] Separability and nonseparability are discussed in Marchand and Russell, and Otto Davis and Andrew Whinston.

[2] In the following, industry A will be the externality-imposing industry. Obviously, it makes no difference which industry imposes the externality; less obviously, it does not matter if each industry imposes an externality because there will be one dominant profit-opportunity locus at each relevant point.

V. Conclusions

Theoretical as well as practical developments in benefit-cost analysis have assumed that marginal cost pricing will be used in conjunction with standard evaluative and investment criteria. However, political and institutional realities often tend to support policies in which public products are priced below marginal cost. In this paper, we have developed evaluative and investment criteria appropriate for the cases in which public products are underpriced (for example, zero priced). Utilization of the new criteria that assume underpricing leads to an increase in efficiency in these situations. Our results are based upon the assumption that consumers will be drawn from a random sample of all consumers who would like to obtain the output at the set price. Even though this assumption is probably the most reasonable and was used in our calculations, others could also have been made. Several alternatives were reviewed to illustrate how they might be incorporated into our analysis. Although we have ignored possible equity benefits that might be obtained by deviating from marginal cost pricing, our analysis does establish the proper measure of efficiency benefits against which equity benefits can be compared in cases of zero or underpricing.

Our analysis suggests that benefit-cost analysis cannot be conducted independently of the pricing policy chosen to allocate an undertaking's capacity. Either prices must be set equal to marginal cost and the standard evaluative criteria used, or if prices are set below marginal cost, then alternate evaluative criteria must be used.

REFERENCES

G. M. Brown, Jr. and M. G. Johnson, "Welfare-Maximizing Price and Output with Stochastic Demand: Reply," *Amer. Econ. Rev.*, Mar. 1973, *63*, 230–31.

C. D. Foster and M. E. Beesley, "Estimating the Social Benefit of Constructing an Underground Railway in London," *J. Royal Statist. Soc.*, Part 1, 1963, *126*, 46–78.

S. H. Hanke, "The Political Economy of Water Resources Development," *Transactions of the Thirty-Eighth North American Wildlife and Natural Resources Conference*, 1973, *38*, 377–89.

J. V. Krutilla, "Efficiency Goals, Market Failure, and the Substitution of Public for Private Action," *The Analysis and Evaluation of Public Expenditures: The PPB System*, Vol. 1, Washington 1969, 277–89.

R. McKean, *Efficiency in Government Through Systems Analysis*, New York 1958.

W. H. Oakland, "Congestion, Public Goods and Welfare," *J. Publ. Econ.*, Nov. 1972, *1*, 339–57.

J. J. Seneca, "The Welfare Effects of Zero Pricing of Public Goods," *Publ. Choice*, Spring 1970, *8*, 101–10.

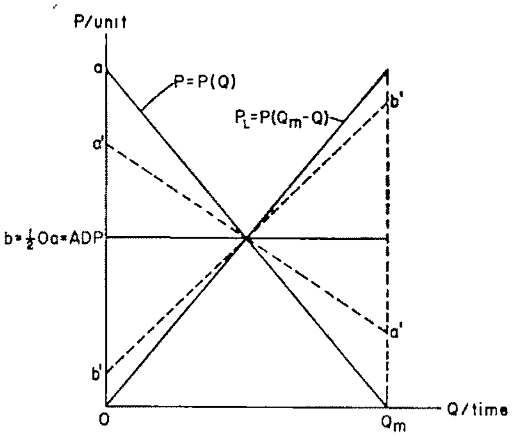
W. S. Vickrey, *Microstatics*, New York 1964.

M. L. Visscher, "Welfare-Maximizing Price and Output with Stochastic Demand: Comment," *Amer. Econ. Rev.*, Mar. 1973, *63*, 224–29.

Figure 3. The marginal benefit is given by the demand-price of the last unit satisfied and is $P_L=P(Q_m-Q_1)$. Since $P(Q)$ is a decreasing function with respect to Q, $P(Q_m-Q)$ is an increasing function with respect to Q (see Figure 4). The three marginal benefit functions associated with the assumptions mentioned in this section are summarized for the case of a linear demand function in Figure 4.

Variations of the above assumptions can also be incorporated into our analysis. In the case where higher willingness-to-pay units have a greater probability (but not equal to one) of being satisfied before lower willingness-to-pay units, the slope of the expected marginal benefit curve will be somewhere between that of $P(Q)$ and ADP, for example, curve $a'a'$ in Figure 4. Alternatively, if lower willingness-to-pay units have a greater probability of being satisfied than higher willingness-to-pay units, the expected marginal benefit curve will have a slope somewhere between that of $P(Q_m-Q)$ and ADP, for example, curve $b'b'$ in Figure 4.

To determine optimum capacity in these cases, one would find the point where the marginal cost curve intersects the appropriate marginal benefit curve from below, and then determine whether or not total benefits are greater than total costs at this point. If total benefits are greater than total costs then this intersection determines the optimum capacity, but if total benefits are less than total costs then the optimum capacity is zero. However, note that all of the marginal benefit curves, except for the demand price curve, have discontinuities at Q_m. These discontinuities must be treated as vertical segments of the marginal benefit curves so that the marginal cost curve passing through the discontinuity represents an intersection of the marginal benefit function from below. This feature of the marginal benefit functions increases the importance of second-order and total conditions in determining the optimum capacity, and gives rise to many situations where the optimum capacity is either zero or Q_m. This is illustrated in Figure 5.



$P=P(Q)$ = highest willingness-to-pay units satisfied first

ADP = each unit has an equal probability of being satisfied

$P_L=P(Q_m-Q)$ = lowest willingness-to-pay units satisfied first

FIGURE 4. MARGINAL BENEFIT FUNCTIONS FOR DIFFERENT RATIONING ASSUMPTIONS

Figure 5 represents a case where the lowest willingness-to-pay units are satisfied first, so the marginal benefit curve is given by the curve labelled P_L. Marginal cost is constant

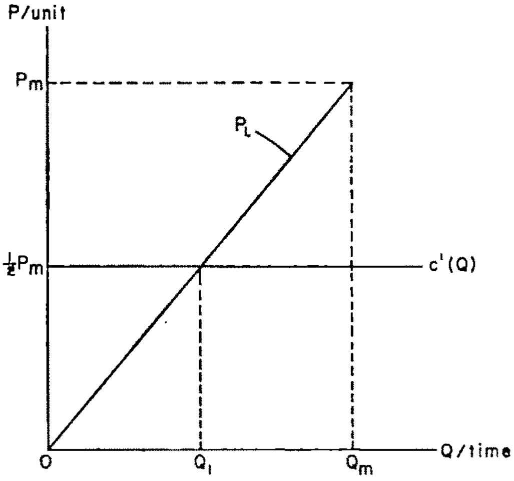


FIGURE 5. CAPACITY DETERMINATION WITH LOWEST WILLINGNESS-TO-PAY UNITS SATISFIED FIRST AND CONSTANT MARGINAL COST

an exclusion cost (*EC*) all nonpayers could be excluded from a project's benefits. If one uses the standard investment criterion for choosing capacity, the incurring of exclusion costs may erroneously appear to be worthwhile in many cases. For example, in Figure 1 capacity Q^* maximizes net benefits when Q goes unpriced, so at zero price Q^* offers greater net benefits than capacity $\hat{Q}$. Therefore, for some values of *EC*, the potential gain in net benefits from excluding nonpayers at capacity $\hat{Q}$ (equal to the difference between area *acd* and area *bcde*) will exceed *EC*. However, if an increase in capacity to Q^* is considered, the potential gain from exclusion (equal to the difference between area *acd* and area *bcf*) will be less than *EC*. Now if we calculate zero pricing benefits at capacity Q^* the pricing policy criterion is as follows: If the gain from exclusion exceeds *EC*, charge marginal cost and choose capacity $\hat{Q}$ (from standard investment criterion); if the gain from exclusion is less than *EC*, price at zero and choose capacity Q^* (from the new zero-pricing investment criterion).

IV. Alternative Nonprice Rationing Assumptions

When effective demand exceeds capacity, it becomes necessary to make some sort of assumption regarding the mechanism used to determine which consumers will be successful in acquiring the output. We believe that the most reasonable assumption, and the assumption employed in our analysis, is one in which each of the demanded consumption units at a given price has an equal probability of being satisfied. Other rationing assumptions, however, are possible. To illustrate how different assumptions would alter our results, we briefly analyze and compare the marginal benefit function for the zero price case under two alternative rationing assumptions: 1) the highest willingness-to-pay consumption units demanded are satisfied first; and 2) the lowest willingness-to-pay units demanded are satisfied first.

In some situations, the first rationing assumption might be the most reasonable. For example, if reservations are used to allocate capacity, it might be reasonable to suppose that the consumers who would be willing to pay the most would be the ones who obtained reservations first or if a lottery in which all consumers had an equal chance of obtaining a reservation were instituted, a perfect speculators market in reservations might develop and the reservations of low-valued consumers would be transferred to high-valued consumers. If queuing and transactions costs were absent in these situations, the result would be the same as if a price were charged to allocate output to its highest willingness-to-pay users. In this case, the marginal benefit function is given by the standard demand-price function $P=P(Q)$, rather than the average demand-price function.

It is also possible to posit situations in which the lowest willingness-to-pay units are satisfied first. Let us assume that the low willingness-to-pay users obtain reservations well in advance of actual consumption and that the high willingness-to-pay users attempt to make reservations on short notice. If no transfers are allowed or if transfer costs are too high, the lowest willingness-to-pay units will obtain reservations first. In this case, the consumption units that are satisfied by any output Q_1 are those from (Q_m-Q_1) to Q_m on the demand-price function. The total gross benefits for output Q_1 are given by the area under the demand-price function $[P=P(Q)]$, from (Q_m-Q_1) to Q_m as shown in

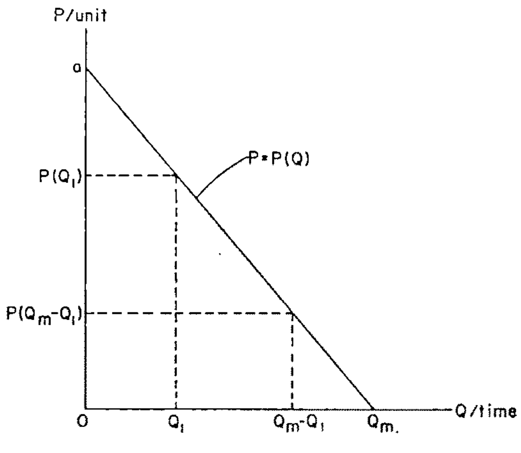


FIGURE 3. VALUATION OF BENEFITS FOR LOW VALUED USERS

This case is analogous to the zero price case. The only difference is that, in the underpricing case, the average demand price is the average over the quantities from zero to $\bar{Q}$, instead of zero to Q_m.

If $\bar{P}=0$, then $\bar{Q}=Q_m$ and equation (6) is identical to equation (4). Now suppose we impose the marginal cost-pricing rule, i.e., $\bar{P}=P(\bar{Q})=C'(\bar{Q})$. This implies that $Q(\bar{P})=\bar{Q}$, which means that Q and $\bar{Q}$, in the first term of equation (6), cancel out and the upper limit of the definite integral becomes $\bar{Q}$. Thus, equation (6) becomes a specific example of the generalized equation (1), and marginal cost pricing and zero pricing are special cases of underpricing. It could be shown, in a manner analogous to that used in the first section of the paper, that welfare losses from using an investment criterion derivable from equation (6) will generally be less than welfare losses resulting from using (2), if a price less than marginal cost is charged.

III. Implications of the Analysis for Project Scale and the Imposition of User Fees

Now that evaluation criteria have been developed, we examine whether the new criteria will yield capacities that are greater than, less than, or equal to those determined by the standard criterion when public products are underpriced. Without information concerning the cost and demand functions, it is impossible to say a priori whether capacity determined by the new criterion is greater than, equal to, or less than capacity determined by the standard criterion. However, if no prices are charged, if increasing costs exist, and if P represents the price indicated by the marginal cost price criterion, the relevant relationships are: if $P=P(Q)=C'(Q)\gtreqless ADP$, then $Q^*\gtreqless\hat{Q}$ where Q^* and $\hat{Q}$ are the capacities determined by the new and standard criteria, respectively. This relationship is shown graphically in Figure 2, which is the same as Figure 1, except that now we have three alternative marginal cost curves. In the case of $C_1'(Q)$, $P(Q)=C_1'(Q)>ADP$ and $Q_1^*<\hat{Q}_1$; for $C_2'(Q)$, $P(Q)=C_2'(Q)=ADP$ and $Q_2^*=\hat{Q}_2$; and for $C_3'(Q)$, $P(Q)=C_3'(Q)<ADP$ and $Q_3^*>\hat{Q}_3$.

The same capacity relationships hold for conditions of constant costs, except when $C'(Q)$ is everywhere equal to ADP; if the new criterion is used, capacity is indeterminate. Furthermore, we know that under conditions of constant costs the new evaluative criterion will determine capacity to be zero, if $C'(Q)$ is greater than ADP, and capacity will be Q_m, if $C(Q)'$ is less than ADP.

Under conditions of decreasing costs, if the total condition (total benefits exceed total costs) is met, the new criterion will always select capacity Q_m, which is greater than the capacity that will be selected by the standard criterion. However, there are cases in which the total condition will be met by the standard welfare measure, but not by the new welfare measure, because total benefits calculated for any given capacity are greater when using the standard measure rather than the new one except at Q_m, where both measures yield the same benefits. In these situations, the standard criterion results in a greater capacity, since there is no investment if the new criterion is used.

Another result of our analysis is a criterion for the choice of pricing policy when the exclusion of nonpayers is possible but costly. High exclusion costs are frequently used to justify the practice of not charging prices for public products. Suppose that by incurring

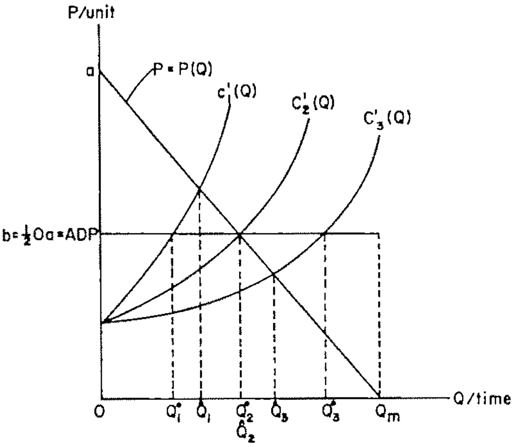


FIGURE 2. CAPACITY DETERMINATION WITH STANDARD AND NEW INVESTMENT CRITERIA ZERO-PRICING CASE

the marginal cost function intersects ADP from above instead of from below,[4] the optimum capacity is Q_m.[5] In this second case, the expected marginal gross benefit function is ADP. Therefore, expected marginal benefits exceed marginal costs until Q_m is reached. At Q_m all demand is satisfied, marginal benefits are zero and the optimum capacity is reached.

We can now see that when prices are not charged for public products, the standard investment criterion (2) and the new investment criterion (5) will not usually lead to the same choice of capacity. It would be strictly coincidental to find $P(Q)=C'(Q)=ADP$, which are the necessary conditions for the same capacity to be selected using both criteria.

The potential welfare gain from the use of equation (5) rather than equation (2) as an investment criterion can be easily illustrated. When it is assumed that price will ration capacity to its highest-valued uses, net benefit is the difference between the area beneath the demand curve and the area beneath the marginal cost curve at the level of Q chosen. Using the standard evaluative criterion of equation (2), optimum capacity in Figure 1 is $\hat{Q}$, where the demand and marginal cost curves intersect. If price $\hat{P}$ equal to marginal cost $C'(\hat{Q})$ is used to allocate $\hat{Q}$, net benefits equal the area *acd*. But if no price is charged, the expected net benefit from capacity $\hat{Q}$ is not the area *acd*. Assuming all consumption units from zero to Q_m have an equal probability of being satisfied, expected net benefit is the area beneath the average demand price curve (here a constant (1/2) Oa since demand is drawn as a linear function) minus the area beneath the marginal cost curve at the level of Q chosen. At capacity $\hat{Q}$, then, expected net benefit at a zero price equals the area *bcde*. Expected net benefit can be increased by using the investment criterion in equation (5) and extending capacity from $\hat{Q}$ to Q^*. At Q^*, average demand price equals marginal cost, and an offer of capacity Q^* rather than $\hat{Q}$ increases welfare by the amount *edf*.

II. The Under-pricing Case

By extending our analysis of welfare measures and investment criteria for the zero-pricing case, we develop a measure of welfare and investment criteria for the more general case of underpricing. In the underpricing case, a positive price $\bar{P}$ is levied for a given output and capacity of $\bar{Q}$, such that $\bar{P}<P(\bar{Q})$.

We assume $Q=Q(P)$ to be the demand function which is the unique inverse of our demand price function $P=P(Q)$. For the underpriced case, where $Q=Q(\bar{P})>\bar{Q}$ and thus greater than capacity, we are again interested in determining the expected value of the demanded consumption units that are satisfied. If we assume each of the $\bar{\bar{Q}}$ units has an equal probability of being satisfied, the expected value (ADP) of a satisfied unit is given by:

$$ADP=\frac{1}{\bar{\bar{Q}}}\int_0^{\bar{\bar{Q}}=Q(\bar{P})} P(Q')dQ',$$

$$\text{for } 0 \leqq Q \leqq \bar{\bar{Q}}$$

This expected value is the average demand price for all consumption units from zero to $\bar{\bar{Q}}$. For any output Q expected total gross benefits B are:

$$B=\frac{Q}{\bar{\bar{Q}}}\int_0^{\bar{\bar{Q}}=Q(\bar{P})} P(Q')dQ'$$

$$\text{for } 0 \leqq Q \leqq \bar{\bar{Q}}$$

For reasons analogous to those given for the zero price case, this benefit function is appropriate only for values of Q that are in the range from zero to $\bar{\bar{Q}}$. Capacity above $\bar{\bar{Q}}$ will not be used because the price $\bar{P}$ will restrict use to $\bar{\bar{Q}}$. The benefit function at this point becomes a constant, $\int_0^Q P(Q')dQ'\cdot\bar{\bar{Q}}$. Subtracting costs $C(Q)$, we obtain the welfare function that we wish to maximize:

$$(6)\quad W(Q)=\frac{Q}{\bar{\bar{Q}}}\int_0^{\bar{\bar{Q}}=Q(\bar{P})} P(Q')dQ'-C(Q)$$

---

[4] Note that the marginal cost function must intersect ADP from below to satisfy second-order conditions for a maximum.

[5] This assumes that the total conditions are met under conditions of decreasing costs.

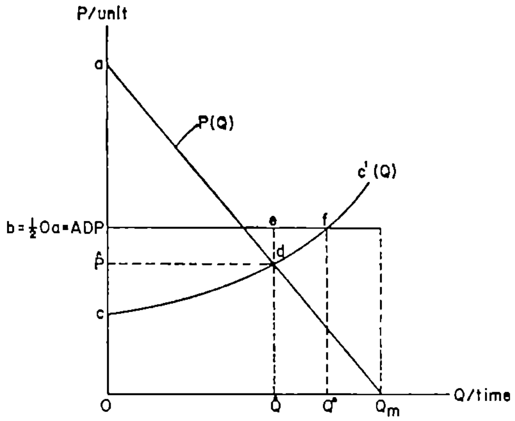


FIGURE 1. CAPACITY DETERMINATION WITH STANDARD AND NEW INVESTMENT CRITERIA
ZERO-PRICING CASE

crats, there is no practical way of allocating capacity to its highest value uses.

In the case where price is not used to allocate capacity, we assume that for any given capacity, each of the demanded consumption units from zero to Q_m has an equal probability of being satisfied. Therefore, the number of demanded consumption units is always greater than capacity, except for the case in which capacity equals Q_m. In this case, each demanded consumption unit has a probability of one that it will be satisfied.[8] Given this assumption, the expected value of a demanded consumption unit that is satisfied is given by the average demand price (*ADP*) for all demanded consumption units (see Figure 1). The average demand price can be written as:

$$ADP = \frac{1}{Q_m}\int_0^{Q_m} P(Q')dQ', \quad \text{for } 0 \leqq Q \leqq Q_m$$

For any given capacity that is less than or equal to Q_m, the expected total benefits B are determined by multiplying *ADP* times the number of demanded consumption units satisfied (capacity). Expected total benefits are represented by:

$$(3) \qquad B = \frac{Q}{Q_m}\int_0^{Q_m} P(Q')dQ' \qquad \text{for } 0 \leqq Q \leqq Q_m$$

Note that the above benefit function is only defined for values of Q in the range from zero to Q_m. For Q greater than or equal to Q_m, all demand is satisfied and there is no rationing problem. Hence, for Q greater than or equal to Q_m, the benefit function is the constant $\int_0^{Q_m} P(Q')dQ'$ and marginal benefits are zero. If we subtract costs $C(Q)$ from equation (3), we obtain the correct welfare measure (expected net benefits) when prices are not used to ration capacity. The new welfare function is represented by the following equation:

$$(4) \qquad W = \frac{Q}{Q_m}\int_0^{Q_m} P(Q')dQ' - C(Q), \qquad \text{for } 0 \leqq Q \leqq Q_m$$

Differentiating equation (4) with respect to Q and setting the results equal to zero gives the first-order condition for maximum expected net benefits:

$$(5) \qquad \frac{1}{Q_m}\int_0^{Q_m} P(Q')dQ' = C'(Q)$$

This condition represents the new investment criterion for this case in which no price is charged for the output from a public undertaking. Optimum capacity occurs at the level where the marginal cost of capacity equals the average demand price.

To complete our discussion of the investment criterion for the zero-pricing case, we must also consider the situations in which the marginal cost function does not intersect the average demand price function in the range of output from zero to Q_m (see Figure 1). First, if the marginal cost function is everywhere greater than *ADP*, the optimum capacity is zero since, for any proposed capacity, the expected benefits are less than costs. Secondly, if the marginal cost function is below *ADP* in the range of output from zero to Q_m or, because of decreasing costs, between evaluation and pricing criteria. The few who recognize the problem prefer to maintain consistency by recommending that standard evaluative criteria be used in conjunction with marginal cost pricing. We believe that the insistence on the use of standard evaluative criteria and marginal cost pricing is unrealistic. Evaluative criteria that are consistent with the zero pricing and underpricing of public products must be developed if efficiency within the public sector is to be improved.

---

[8] This assumption seems most appropriate, unless there is a definite reason for suspecting that the majority of output will go to either high willingness-to-pay users or to low willingness-to-pay users (see Gardner Brown and M. Bruce Johnson and Michael Visscher). The effect of various rationing assumptions upon our analysis will be discussed later in the paper.

In this paper, we reformulate the zero-price evaluative criterion developed by Seneca and build upon it to derive appropriate investment criteria. We then extend the analysis to the general case of underpricing, which was touched upon by Vickrey. Welfare levels for zero pricing and underpricing are compared for standard and reformulated investment criteria. Equity effects are not considered.[1] The policy implications derived from our analysis are also presented.

I. The Zero-Pricing Case

We begin our analysis with a discussion of the limiting case in which a public product's price is zero. For the output of the project in question, we have an aggregate demand price function $P=P(Q)$, where $dP/dQ<0$, $P(0)$ is finite, and there is a finite output Q_m, where $P(Q_m)=0$. If prices are used to allocate capacity, it is assumed that the definite integral of the demand price function is an adequate measure of total benefits. We further assume that operating costs are zero, and capacity is measured in terms of possible output Q for a given project scale. Hence, we have a capacity cost function represented by $C=C(Q)$. Our analysis is simplified by focusing on a single time period in which capacity depreciates instantaneously at the end of the demand period (for example, one year) for which our demand price function is relevant. Lastly, we assume that congestion costs do not exist.

Our assumption regarding congestion costs indicates that there is a direct relation between the number of production and consumption units.[2] This assumption excludes the possibility that for a given number of production units there exist alternative consumption units of varying quality, where the quality is a function of use. Our model, therefore, is developed for the case in which there is a one-to-one correspondence between output and consumption. For example, if an irrigation project is being considered, the amount of water consumed cannot be greater or less than the net amount (gross water diverted less evaporation and leakage) of water produced. Our assumption implies that, regardless of the method used to allocate the output, individuals' valuations of the output will not be altered. This is necessary to maintain independence between the demand function and the price charged.

The problem for project planners is to determine the level of output ($\hat{Q}$) (capacity) that maximizes the appropriate welfare (W) function. The welfare function used in standard analysis is represented by the following equation:

$$W(Q) = \int_0^Q P(Q')dQ' - C(Q) \tag{1}$$

To maximize welfare, we differentiate equation (1) with respect to Q and set the results equal to zero, obtaining the first-order condition:

$$P(Q) = C'(Q) \tag{2}$$

This condition represents the standard investment criterion; capacity should be chosen so that the value of the marginal unit of capacity equals its marginal cost. However, this investment criterion leads to a welfare maximum only if capacity is allocated so that only points on the demand price function with values greater than or equal to $P(\hat{Q})$ are served. The familiar marginal-cost pricing rule in which a price of $\hat{P}=P(\hat{Q})=C'(\hat{Q})$ is charged would lead to a welfare maximum, since users who do not value capacity at levels equal to or greater than the price $\hat{P}$ would be excluded. But many times no price is charged for public products and, lacking omniscient bureau-

---

[1] For a discussion of equity arguments for and against the underpricing of public products, see Krutilla.

[2] This relation is discussed by William Oakland.

Public Investment Criteria for Underpriced Public Products

By GENE E. MUMY AND STEVE H. HANKE*

Procedures for the design and operation of public undertakings have traditionally separated the evaluation of project benefits and the pricing of project outputs. Benefit-cost calculations have been based upon standard efficiency criteria without regard to the pricing policy utilized during the operating phase. An undertaking, whose scale has been appropriately determined by equating incremental benefits and costs, is deemed worthy of investment if its total benefits exceed costs. However, if the pricing policy that governs the sale of output is not explicitly taken into account in the benefit-cost analysis, serious errors in the estimation of benefits and, hence, the scaling of facilities will be encountered.

In practice, no widespread recognition of the effects of ultimate public project pricing policies on prior benefit calculations is evident (see Hanke). For example, the Water Resources Council has recently provided new rules for benefit-cost analysis that are far more comprehensive than any previous federal effort, but the new standards do not mention pricing policies or how they might alter the performance of public undertakings.

In the academic literature, very few have noted the dilemma created by not explicitly taking into consideration pricing policies when conducting benefit-cost analyses. John Krutilla is one of the few economists to explicitly recognize the general implications of separating evaluation and pricing procedures. In their classic benefit-cost study of the Victoria Line, Christopher Foster and Michael E. Beesley assumed that the fares imposed by the Victoria Line would make no difference in the calculated net rate of return for the project. Both Christopher Winsten and Alan Walters, in discussing the Foster-Beesley paper, correctly criticized this assumption. They argued that net benefits will be reduced by charging underground fares that exceed the appropriate marginal cost, adjusted for nonoptimal prices that existed in competing transport modes. In order to maintain consistency between evaluation and pricing criteria, Krutilla, Winsten, and Walters have recommended that standard evaluative criteria should be used in conjunction with marginal cost pricing. Roland McKean also expressed a preference for the use of standard evaluative criteria and marginal cost pricing. McKean realized, however, that prices would be set below marginal costs in some cases. In these situations, he suggested that project benefits be conservatively estimated. One of the few economists who has analytically dealt with the problem of evaluative criteria modification in the face of prices that are below marginal cost is William Vickrey. In a discussion of peak-load pricing, he recognizes that there are institutional, technical, and economic reasons why prices cannot always be set equal to the peak-load marginal cost. To maintain consistency between pricing and evaluation, Vickrey develops benefit measures for various types of nonprice rationing assumptions. Joseph Seneca has also dealt analytically with the problem of separating pricing and evaluation. He developed a measure of benefits for a given output in the extreme case in which no price is charged.

In our reading of the benefit-cost literature, we have found that most analysts ignore the problem of maintaining consistency variation measure entails a higher resultant level of welfare than the compensating variation," and that the " . . . (money value) of any given benefit to the consumer is greater for a higher level of welfare than it is for a lower." On the first count, since both measures are usually defined in relation to the same price change they can also only entail a similar welfare change, whereas the second matter appears to be a statement in sharp contradiction to Mishan's own point concerning the possible behavior of the MUY.[3]

---

* Assistant professor, Virginia Polytechnic Institute and State University, and associate professor, The Johns Hopkins University and University of California, Berkeley, respectively. This paper was prepared with the assistance of a research grant from the National Science Foundation, NSF-RANN, GI-34869. We wish to thank John Boland, George Borts, Gardner Brown, Paul Bugg, Anthony Fisher, Jack Knetsch, John Krutilla, Joseph Seneca, Roger Sherman, Ralph Turvey, Michael Visscher, Richard Walker, and an anonymous referee for their helpful suggestions.

Again Mishan is misleading when he states that "In the ordinal system, therefore, these two measures of the consumer's gain are *not* approximations to a more precise measure." I have already suggested that this concern with ordinality may be misplaced, but in any case all that I explained in my original paper was that neither measure corresponded exactly to the effect that Hicks and Marshall appeared to wish to evaluate, the money value of a utility change associated with a price change.[4] I would therefore hold to my original view in this matter and also in respect of Mishan's final point concerning my suggestion as to the evaluation of the effects of multiple price changes. Particularly in cases where there are both price rises and falls my suggested path of integration increases the likelihood of obtaining a measure somewhere between the (generalized) compensating and equivalent variations.

---

[3] Alternatively Mishan may mean here (but does not say) that a greater welfare increase from a given initial position will yield a greater money value of the benefits. That this is incorrect can be seen directly from the fact that the compensating and equivalent variations are just two possible different money evaluations of a given welfare change from a fixed initial position. By varying the path of integration almost any money value may be placed on any welfare change.

[4] Such a change is clearly defined by a path of integration between A and B along which income is held constant and only the price of X changes.

REFERENCES

M. E. Burns, "A Note on the Concept Measurement of Consumer's Surplus," *Amer. Econ. Rev.*, June 1973, *63*, 335–44.

A. E. Danese, *Advanced Calculus*, Vol. 1, Boston 1965, 98–171.

S. Ichimura, "A Note on the Concepts of Consumer's Surplus," *Econometrica*,, July 1953, *21*, 484–85.

E. J. Mishan, *Cost-Benefit Analysis*, London 1971, 328–29.

———, The Concept and Measure of Consumer's Surplus: Comment," *Amer. Econ. Rev.*, September 1975, *65*, 708–09.

The Concept and Measure of Consumer's Surplus: Reply

By MICHAEL E. BURNS*

I am grateful to E. J. Mishan for drawing attention to a minor error in my recent paper, but I am unable to agree with him either as to the significance of the stated imprecision or in respect of certain other observations that he makes. The force of Mishan's comment concerns my discussion of the compensating and equivalent variations and I shall deal with this matter first, in particular with his claim that this discussion is not valid within a Hicks-Allen ordinal utility system.

My approach was first to describe the measures concerned, reinforcing this description by an illustration already widely used in the related literature, my Figure 1.[1] I proceeded to explain that the divergence between these measures reflected only that "... any evaluation of the income equivalent will depend upon the value of the marginal utility of (real) income (*MUY*) along the path on which the evaluation takes place" (p. 339). This statement draws only on the theory of line integrals as available in most advanced calculus texts, and is of course of general validity irrespective of whether one considers either a two-good or an *n*-good world and also quite independently of the behavior of the *MUY* as we shift from one utility level to another. Again, my derivation of the actual magnitudes of these two Hicksian measures that follows from the path-dependent characteristics of line integrals is quite general and completely applicable to an ordinal utility system.[2] Thus Mishan's criticism appears to be justified only in the specific case where for interpretative purposes I translated the results onto Figure 1 in a manner that precluded the possibility of the *MUY* increasing when income alone was increased. This point is well taken and the paper stands amended in this respect, but its importance to the exposition as a whole should not be overrated; indeed the sentence concerning the relative values taken by the *MUY* at positions *A*, *B*, *C*, and *D* in Figure 1 could have been omitted without affecting the content of the paper.

As it is, Mishan offers a correction concerning the ordinal interpretation of Figure 1 that I am unable to accept. Thus if the good *X* is normal in the (simply connected) region *ACBD*, and bearing in mind that we are only considering either *changes in income and/or in the price of X*, I would suggest that the *MUY* at *B* is greater than at *D* (and greater at *C* than *A*). Were the *MUY* to change in a contrary direction along an indifference curve (i.e., increase while moving from *B* to *D* or from *C* to *A*) it would always be possible to find a subregion of *ACBD* wherein *X* is an inferior good. Short of rejecting the notion that normality should be defined within a simply connected region, it would seem that Mishan's interpretation is incorrect and that my original statement needs less modification than might have appeared to be the case.

Mishan raises a number of other points that I do not find convincing. In his further discussion of Figure 1 he states that "for a fall, say, in the price of *X*, the equivalent that the difference in the Y-measures of compensating variation and equivalent variation for a change in the price of X is to be explained in terms of an income effect expressed in Y units. For a fall, say, in the price of X, the equivalent variation measure entails a higher resultant level of welfare than the compensating variation. And the Y-value (or "money" value) of any given benefit to the consumer is greater for a higher level of welfare than it is for a lower.[4]

---

* University of Manchester.

[1] Thus when Mishan refers to "... Burns' two-good system ..." he is actually referring to an illustrative device made use of by many others, including himself (1971), p. 329.

[2] A. E. Danese is one of many texts dealing with the theory of line integrals, and indeed since writing my original paper I have discovered that Shinichi Ichimura, pp. 484–85, had pointed to the relevance of such theory to this identical problem over twenty years ago. On the question of ordinal utility, the evaluation of any of the line integrals concerned requires only information of the values of x and dp along specific paths, some of these paths being indifference curves. These evaluations are therefore independent of actual utility levels, depending at most on the loci of the indifference curves.

In the ordinal system, therefore, these two measures of the consumer's gain are *not* approximations to a more precise measure. Each itself is a precise measure. If when the price of X falls, the buyer is willing to pay $10 at most in order to avail himself of the lower price, this $10 is exactly what the fall in price is worth to him. As for the equivalent variation, if the buyer were denied the opportunity of buying X at the now lower price, a sum of, say, $12 would be such as to make him just as well off as he would have been if, instead, he had been given the opportunity to buy X at the lower price.

Each measure then is a precise answer to a different, though potentially relevant question. As such, each has a clear operational value. Moreover, as Hicks and others have shown, it is the individual's demand curve that is an approximation to these two precise measures.[5]

---

[4] This is true whether the individual is responding to changes in prices as a consumer or as a producer, or as both simultaneously, and therefore whether the two measures of such changes in his level of welfare are appropriately referred to as consumer's surplus or as worker's rent or both, respectively. See the author.

[5] Accepting the area under the individual's demand curve as a reasonable approximation to the precise measures of consumer's surplus, Burns' measure of consumer's surplus of a simultaneous fall in the price of two goods i and j does not accord with that which follows from Hicks' 1956 proposal. Whether the two goods are substitutes or (as implied by the construction of Burns' Figure 2) complements, the measure of consumer's surplus is obtained by treating a simultaneous fall in prices as if they occurred in sequence, paying close attention to the *ceteris paribus*.

REFERENCES

R. G. D. Allen and J. R. Hicks, "A Reconsideration of the Theory of Value," *Economica*, Part I, Feb. 1934, *1*, 52–76; Part II, May 1934, *1*, 196–219.

M. E. Burns, "A Note on the Concept and Measure of Consumer's Surplus," *Amer. Econ. Rev.*, June 1973, *63*, 335–44.

J. R. Hicks, *Value and Capital*, 2d ed., Oxford 1956.

———, *A Revision of Demand Theory*, Oxford 1956.

E. J. Mishan, "Rent as a Measure of Welfare Change," *Amer. Econ. Rev.*, June 1959, *49*, 386–95.

The Concept and Measure of Consumer's Surplus: Comment

By E. J. MISHAN*

In his recent note in this *Review*, Michael Burns' equations are correct, and his exposition in parts may be satisfactory as an interpretation of a Marshallian model of consumer's surplus, one in which the marginal utility of real income and of each good diminishes. In such a model an exact measure of consumer's surplus has to be exact in "real" terms, or in "utils." And in order to express each "util" of consumer's surplus unamibiguously as a sum of money, the marginal utility of money[1] must indeed remain constant despite any change in price that generates the consumer's surplus. In so far, then, as the marginal utility of money does not remain constant with respect to the price change in question—and, of course, it cannot do so with diminishing marginal utility of real income—any money measure of consumer's surplus derived on the (false) premise that it does remain constant can, at best, be only an approximation to some "true" util measure.

Burns' exposition cannot, however, be accepted as an interpretation of the Hicks-Allen approach to consumer theory. In their system, utility is ordinal and the marginal utility of real income, identified as the Lagrangian multiplier λ, can be constant, increasing, or decreasing, as indeed can the marginal utility of any of the goods. Thus the utility "heights" of a succession of evenly spaced indifference curves can increase or diminish in steepness[2]—the utility surface so generated being either concave, plane, or convex according as the marginal utility of real income is increasing, constant, or diminishing. But, as we all know, we do not need this utility axis in the Hicks-Allen system: we derive all the necessary theorems from the shape of the indifference curves, which shape tells us nothing at all about absolute utility.

It follows that Burns' explanation of the difference in the Y measure of compensating and equivalent variations for "normal" goods[3] is not valid within a Hicks-Allen ordinal utility system. And if we happen to have *increasing* marginal utility as total utility increases, in particular when we move from I_1 to I_2 in his Figure 1, it does not necessarily follow that the marginal utility of real income is greater at C than it is at A or at B (or less at D than it is at A or at B). All that can be said from the ordinal interpretation of the figure is that the marginal utility of Y *relative* to the marginal utility of X diminishes as Y is increased with X remaining constant (though, again, the *absolute* marginal utility of Y itself need not diminish; it can increase). And this implies that the rate of substitution $\partial Y/\partial X$ increases as Y is increased with constant X. But this is a corollary of good X being a "normal" good, defined as one for which more is bought as the consumer's budget increases.

It follows from the construction of Figure 1, in which X is depicted as a normal good, empirical results in my paper. If variables in nominal terms have to be used in the consumption function, it is essential that all variables be measured in the same units, and from this standpoint, the definition as well as the measures of accrued gains employed in my paper are correct. Accrued gains, like income and consumption, are a *flow*, and under my definition, they are expressed in the same units as these other variables. M-P's proposed adjustment to eliminate my "mismeasurement" therefore is unnecessary.[4]

---

* London School of Economics. I am grateful to Sir Roy Allen for glancing through this comment and for drawing my attention to a minor error.

[1] In an n-good cardinal utility system in which the money income $M = \sum p_i x_i$, the Lagrange multiplier λ is equal to $\partial U/\partial M$ or the marginal utility of money income. However, if a single good, say x_1, is taken as numeraire and priced at unity, the marginal utility of this good x_1 is also equal to λ (since $\partial U/\partial x_1 = \lambda p_1 = \lambda$). Thus given Burns' two-good system, if good Y is the numeraire and priced at unity, the marginal utility of Y is also equal to λ. If Y is interpreted, instead, as "all goods other than X at fixed prices," and is also regarded as numeraire equal to unity, it is necessary that the purchase price of a unit of each of these goods (other than X) add up to unity.

[2] As clearly indicated by Hicks (1946), pp. 13–16.

[3] Apparently Burns' interpretation of a normal good is one for which any increase in quantity entails diminishing marginal utility. But a normal good on the Hicks-Allen definition is, of course, consistent with marginal utility that increases, diminishes, or remains constant.

In the light of the discussion above, M-P's argument that the significant coefficient on G_t reported in my paper *could* result from the nominal variables used in empirical estimation is mere conjecture which they should verify, especially because they seem to be claiming that they have the "appropriate" data.

III. The Empirical Results

M-P are also skeptical of my assertion that the empirical results in the wealth approach are much sharper because it is ". . . based primarily on a comparison of the Durbin-Watson statistics" (p. 700, fn. 1). In my paper I said:

> The results are much sharper than in the income approach: the Durbin-Watson statistic suggests serial independence of residuals in all cases, W_t^* and Y_t^* are highly significant, the coefficients have the expected signs (positive), and their magnitudes are not affected much by whether accrued or realized gains are included in the equation. The normal rate of return on wealth ($\hat{\beta}=k\beta/k$) is estimated between .06 and .08, which seems quite reasonable. All in all, equation (6) appears to be much better specified than equation (4). [p. 875]

M-P's statement, to say the least, is a grossly unfair representation of my position. In response to their point that the Durbin-Watson statistic might be biased because of a lagged dependent variable in the consumption function, I present below the Cochrane-Orcutt estimates of equation (6) for accrued gains (see fn. 18, p. 873 of the original paper).

$$\underset{(-0.34)}{C_t = -\,3.052} + \underset{(6.17)}{0.790Y_t^w} + \underset{(2.89)}{0.051W_t^*}$$

$$SEE = 1.96, \quad R^2 = 0.999, \quad D.W. = 2.17, \quad \hat{\rho} = \underset{(-0.11)}{-\,0.031}$$

$$C_t = \underset{(21.87)}{0.748Y_t^w} + \underset{(9.27)}{0.057W_t^*}$$

$$SEE = 1.78, \quad R^2 = 0.999, \quad D.W. = 2.20, \quad \hat{\rho} = \underset{(-0.22)}{-\,0.063}$$

where *t*-values are enclosed in parentheses. It is obvious that these results are almost the same as those reported in my paper. Also note that $\hat{\rho}$ is not significantly different from zero.

REFERENCES

K. B. Bhatia, "Capital Gains and the Aggregate Consumption Function," *Amer. Econ. Rev.*, Dec. 1972, *62*, 866–79.

———, "The Estimation of Accrued Capital Gains on Individuals' Corporate Stock Holdings," *J. Amer. Statist. Assn.*, Mar. 1973, *63*, 55–62.

M. B. McElroy and J. C. Poindexter, "Capital Gains and the Aggregate Consumption Function: Comment," *Amer. Econ. Rev.*, June 1975, *65*, 700–03.

---

nonfarm real estate were estimated by subtracting net capital formation from change in net worth during a year ($W_t - W_{t-1}$), but the procedure for estimating gains on corporate stock was different and more complicated. It involved the use of a stock price index and a series of adjustments to correct for possible errors in the data series. See the author (1973) for details.

[4] M-P's proposed measure is

$$G_t' = W_t - \left(\frac{P_t}{P_{t-1}}\right) W_{t-1} - \left(\frac{P_t}{\overline{P}_t}\right) I_t$$

It will, in effect, express accrued gains in prices prevailing at a *point* in time, the end-of-year. But the other flow variables in the consumption function, namely C and Y, are expressed in prices prevailing at *different points* during the year: these are simply totals of quarterly or monthly data without any adjustments for changes in price level. For example, the national accounts estimates of personal income in current prices are $Y_t = Y_{t1} + Y_{t2} + Y_{t3} + Y_{t4}$, and not

$$Y_{t1}\left(\frac{P_4}{P_1}\right) + Y_{t2}\left(\frac{P_4}{P_2}\right) + Y_{t3}\left(\frac{P_4}{P_3}\right) + Y_{t4}$$

as consistency with the M-P measure of accrued gains would require.

Clearly, our reference must be to savings and dissavings realized for the purpose of shifting expected consumption over time, and we must seek the best such measure derivable from the data at hand. Given the *BLS* breakdown of expenditures, that measure excludes gifts or contributions and personal-insurance premiums, even though part of the latter are true savings which, ideally, we would want to retain. The matter of whether these expenditures should be added to consumption (as done by Thurow under his second option) or subtracted from income is to be resolved empirically: Since the insurance /contributions total varies roughly in proportion with after-tax income, but not so with consumption, the subtraction from income is more consistent with our specification of the analytic model.

Accordingly, the empirical work in Section IVD was redone, using after-tax income net of personal-insurance premiums and contributions. The income measure was further corrected by addition of "other money receipts," a category left out of the *BLS* definition of money income.[21] The results are similar to those obtained in *D*, although the maximum excess of desired over actual consumption occurs even later in life and is of lesser magnitude than under the initial definition of income.[22] This confirms the suspicion that, for the average-income household, ***institutional obstacles to additional borrowing are experienced in the years when the cost of supporting and educating dependents is at its highest, and when expensive comforts have become an important condition for the enjoyment of life.***

F

The effect of two biasing factors identified in the course of the analysis accounts for slight inconsistencies observed in the estimates: (a) a desired consumption rising above actual in the oldest age class (under the adjusted estimate of Section IVE only), and (b) a desired consumption slightly under actual in the youngest age class.

In general, the existence of any discrepancy between desired consumption d_A and intended consumption $\hat{c}_A$ at the zero-saving observation for age A, generates an estimate $\tilde{g}_A$ in (29′) that deviates from its true value $g(A)$ by:

$$\tilde{g}_A - g(A) =$$

$$\Delta_A \frac{\theta_A(C_A)}{(1-p)+p[y_A^* + r(A)]/[F(A)+r(A)]}$$

where Δ_A is the discrepancy $(\hat{c}_A - d_A)$ and $\theta_A(C_A)$ is the value of the biasing function introduced in (16), both at the zero-saving observation.[23] The ratio multiplying Δ_A does not deviate much from 1, since the denominator has an empirical range of 0.9 to 1.5 and $\theta_A(C_A)$ can be assumed close to 1. Roughly speaking, then, the bias of the estimate $\tilde{g}_A$ is of the same order as the discrepancy Δ_A.

(a) It was shown in Section II that we generally have $\hat{c}_A > d_A$, or $\Delta_A > 0$ at actual zero-saving points, meaning that the estimated profile of desired consumption is biased upward. This bias would be especially strong for later-life estimates, where zero saving occurs in the lower range of income profiles after a long history of constrained dissaving.

(b) A reverse (and likely larger) bias affects estimates in the youngest age groups, where zero-saving observations occur in the high range of income profiles. As pointed out in Section IIID, zero-saving *observations* imply an intended dissaving if the income profile is high, and an intended positive saving if the profile is low. The implication in the first case is that desired consumption is above actual, i.e., $d_A > \hat{c}_A$, or $\Delta_A < 0$, as long, that is, as we can assume the discrepancy thus created to exceed that identified above. This suggests a downward bias of the estimated profile of desired consumption for the youngest age classes.

---

[21] "Other money receipts" include inheritances and occasional large gifts of money less taxes and associated expenses; and net receipts from the lump-sum settlement of fire and accident insurance policies.

[22] Zero-saving incomes (under the new income definition) were obtained through careful interpolations of available class averages. An ambiguity occurs in the "under-25" age interval, where savings fluctuate between positive and negative values over four income classes. Fortunately, the estimate of desired consumption for that age interval is relatively insensitive to variations of the zero-saving income in its possible range. Another difficulty occurred in the estimation of $r(A)$, which was computed as the average of point estimates for the 10th percentile and 90th percentile income profiles only. Point estimates for the 50th percentile were eliminated because of their excessive sensitivity to observational error (the denominator $W^i - 1$ in the estimation formula (fn. 19) is a low $-.086$).

[23] The expression for $g(A)$ is easily obtained by repeating Section IVD's computations with $\hat{c}_A = d_A + \Delta_A$.

It is not unrealistic to expect that, in the middle range of ages, the two identified biases approximately cancel one another, so that the estimated profile of desired-consumption can be accepted with a reasonable degree of confidence.[24] Nevertheless, the conclusions are in need of testing against more direct evidence of individual saving behavior under actual or experimental variations in household access to consumption loans.

---

[24] In common with the work of Thurow and his reviewers, no effort was made to adjust cross-sectional data on income and consumption by age interval for average income growth over time. This is not likely to affect the measure of discrepancy exhibited in Figure 6, but it does mean that both the desired and the actual consumption profiles are, in fact, tilted upward on the upper-age side.

REFERENCES

R. W. Hartman, *Credit for College,* New York 1971.

D. B. Johnstone, *New Patterns for College Lending,* New York 1972.

B. Motley and S. A. Morley, "The Optimum Lifetime Distribution of Consumption Expenditures: Comment," *Amer. Econ. Rev.,* Sept. 1970, *60,* 738–42.

L. C. Thurow, "The Optimum Lifetime Distribution of Consumption Expenditures," *Amer. Econ. Rev.,* June 1969, *59,* 324–30.

———, "The Optimum Lifetime Distribution of Consumption Expenditures: Reply," *Amer. Econ. Rev.,* Sept. 1970, *60,* 744–45.

———, "Erratum," *Amer. Econ. Rev.,* Mar. 1971, *61,* 248–49.

K. H. Young, "The Optimum Lifetime Distribution of Consumption Expenditures: Comment," *Amer. Econ. Rev.,* Sept. 1970, *60,* 736–37.

U.S. Bureau of Labor Statistics *(BLS), Survey of Consumer Expenditures, 1960–1961, Consumer Expenditures and Income Cross-Classification of Family Characteristics, Urban United States,* Supp. 2, Part A to *BLS* Report 273–38, July 1964.

The Optimum Lifetime Distribution of Consumption Expenditures: Reply

By LESTER C. THUROW*

There are several levels upon which a reply might be written to Andre Daniere. Given what I presume is a high decay rate in the reader's interest in comments and replies let me start with the final results and work backwards. Qualitatively we agree. Both Daniere's methods and my methods lead to the need for some redistribution of lifetime consumption expenditures toward younger ages. Daniere says that we disagree quantitatively ("the gap is small in relation to that calculated by Thurow"). In my original article the optimum redistributions were $31,802 under the first definition of savings and $21,618 under the second definition of savings out of an expected lifetime income of $342,246. Since Daniere does not choose to tell us the size of his optimum redistributions, I do not know whether his quantitative estimates lie outside my range of Baysian reasonableness. Eyeball integration of Figure 6 (his final results) leads me to the conclusion that he also has a substantial amount of redistribution.

Perhaps even more to the point, the net result of all of Daniere's modifications of my original estimation technique is to give a result that is, as he admits absurd (see Section IVE). He states, "One disturbing feature of the desired-consumption profile estimated in the previous section is that it exceeds actual consumption over a range (age 25–45) in which income is itself higher than consumption: If the average-income household wishes to consume more in that range, why does it not do so by spending more of its income?" Why indeed? There is no reason for actual consumption to fall below desired consumption if desired consumption is itself less than actual income. He does not even try to explain the even greater absurdity of estimating desired consumption to be less than actual consumption in the earliest age ranges (see Figure 6).

To escape from his embarrassment of having estimated results that are logically impossible, he suddenly decides to shift from my first definition of savings to a strange transformation of my second definition of savings. In my original article zero savings was defined in two ways. 1) Zero savings occurs when after-tax money income is equal to expenditures for current consumption, and 2) zero savings occurs when after-tax money income is equal to expenditures for current consumption plus gifts, contributions, and personal-insurance payments. In the second case gifts, contributions, and insurance payments are considered consumption expenditures rather than savings. Daniere, however, wants to subtract gifts, contributions, and insurance payments from income rather than add them to current consumption expenditures. (Which you do makes a difference since estimated lifetime consumption expenditures must be adjusted up or down (see fn. 14 in my original article) to be brought into equality with actual lifetime incomes.)

Gifts, contributions, and insurance payments might be considered savings or they might be considered consumption, but they clearly are not vanishing income. Why then does Daniere want to treat them as vanishing income? His own justification does not make any sense. "The matter of whether these expenditures should be added to consumption (as done by Thurow under his second option) or subtracted from income is to be resolved empirically: Since the insurance/contributions total varies roughly in proportion with after-tax income, but not so with consumption, the subtraction from income is more consistent with our specification of the analytical model." What this means I leave for others to decipher. In my understanding of economic accounting there are savings or investment expenditures and consumption expenditures. Income is by definition equal to the sum of the two. There is nothing called vanishing income.

---

* Massachusetts Institute of Technology.

Why then this peculiar invention of vanishing income? I know of only two possible explanations. To treat gifts, contributions, and insurance in the normal way would be to obtain another absurd result or an estimate closer to my original results.

Given what are admitted to be silly results, the only real questions revolve around which of his three basic adjustments (tilt, permanent income, and the inequality between lifetime income and consumption) causes the absurd results.

The easiest of these to deal with is tilt. As Daniere correctly points out, any restrictions on actual dissaving in early years will raise observed consumption in later years as individuals attempt to compensate for their lack of consumption in early years (see his Figure 2). This will distort the calculated relationships between actual consumption and desired consumption. What Daniere neglects to point out, however, is the kind of bias introduced by a failure to allow for the current impact of past disequilibria. Early constraints on dissavings will raise consumption levels in middle and later years and thus my estimates will overestimate—not underestimate—the amount of consumption that the family would ideally like to do in its middle and latter years. Thus a correction for tilt *must* increase the amount of desired redistributions.

How then can you square the known direction of the bias with the empirical measures of the bias at the end of his Section IVD? I can only assume that Daniere made an algebraic error and reversed the signs so that we made exactly the opposite correction for those that should have been made.

But there is something else wrong with Daniere's tilt calculations. For some unknown reason he assumes that the bias is zero at 50 years of age. The bias could be zero at the initial age or at the final age, but it cannot be zero at a middle age. At the initial age earlier disequilibria cannot affect consumption since there are no earlier disequilibria by definition. At the final age disequilibrium can have no impact since the individual has had his entire lifetime to correct for the early disequilibria. Thus his tilt corrections are wrong with respect to both sign and lifetime pattern.

Daniere's second supposed correction is for neglecting the difference between permanent and transitory income. Here it is not clear whether Daniere is working with a permanent income model where the expected lifetime value of transitory income is zero or where the expected lifetime value of transitory income is positive. In the second case lifetime optimization is clearly a more difficult problem than the one I proposed since the individual does not know his or her lifetime income. Obviously you cannot optimize vis-à-vis an unknown constraint. If individuals systematically overestimate lifetime incomes, early consumption expenditures will be too high relative to later consumption expenditures and if individuals systematically underestimate lifetime income, early expenditures will be too low. As a result my estimates are only biased toward too much redistribution if you believe that individuals systematically overestimate lifetime incomes. If lifetime transitory incomes are positive, then my estimates are too low, not too high.

But let's go back to the hypothesis that the expected lifetime value of transitory income is zero. In this case there should be no bias in my results. The distinction between permanent and transitory income leads to conclusion that there will be too many individuals in the upper or lower tails of the distribution at each age but it does not affect the mean of the distribution. Since I was working in the middle ranges of the income distribution and with groups of individuals, no bias is introduced. The bias would be important if I were working with individual data, but I am not. Presumably those that are zero savers by accident (transitory income placed them in a nonoptimal consumption class) cancel each other out and the observed income of the class of zero savers has not been biased by the existence of these accidental zero savers.

The third correction is designed to take account of the "fact" that lifetime consumption does not equal lifetime income for high income classes. Several observations need to be made. First the observation is deduced from the fact that the 90th percentile of the income distribution has a lifetime pattern of consumption that is always below its income (see Figure 3). Daniere forgets his permanent-transitory distinction and the easy explanation that the 90th percentile is not necessarily the same families. Second, bequests at death can be regarded as a form of consumption so that individuals desire to consume a lot at the point of death. If this is true, the appropriate correction is not to raise desired consumption during the middle years, but to raise consumption at death. Assuming a positive rate of time preference this would lead to more saving during the middle years to increase consumption at death. Once again the failure to allow for this possibility means that my estimates are biased toward too much consumption in the middle years, not too little. Third, the middle income family—the family that I am concentrating upon—does not leave a substantial amount of assets at death. Its lifetime consumption does seem to be near its lifetime income and therefore a correction is not necessary for a middle income family.

I also agree with Daniere that there is a problem with my estimation technique if you assume that the basic lifetime pattern of tastes changes as lifetime incomes rise or fall. This is a possibility that was discussed in the original article and to some extent investigated empirically (see fn. 14 in my article).

Where does this leave us? Since my original article rested on some strong assumptions, I would be the first person to acknowledge a better calculation that succeeded in lifting some of those assumptions. Unfortunately Daniere's calculations are not an improvement.

A Model of a Highway System Financed by a Trust Fund

By E. G. DAVIS AND J. E. OLSON*

> What made the road in general was man's necessity of passing from one point to another; but what made the Highway was the Vehicle.
>
> It was when man began to extend his carrying capacity (and perhaps his speed) by the creation of the Vehicle that the road took on a special form and became what it has remained among us: the Highway. The cause determines the effect, but, what is more, the effect is contained in the cause.
>
> [*Hilaire Belloc*, p. 1]

Mutual interdependence between the vehicle and the highway did not stop with their creation; in the United States, this interdependence has been institutionalized by legislative action. Highway construction and maintenance are financed by vehicle-related taxes levied at both federal and state levels which are, in the main, dedicated to these purposes by law.[1] Chief among these is the tax on gasoline.[2]

There is, in addition, a behavioral relationship between the vehicle and the highway: vehicle usage is presumably affected by the size and quality of the existing highway system. Moreover, revenue from gasoline taxes would be determined primarily by vehicle usage. The fact that the revenues are dedicated to highway purposes by law completes what is, in essence, a dynamic process.

In this paper we present a simple model of the dynamic relationship between highway size and vehicle usage and attempt to estimate its parameters.

I. The Model

Our model of the process of highway construction and maintenance differs from reality in two major respects: we assume one national trust fund rather than separate state and federal funds, and we assume the only source of revenue for the fund is the tax on gasoline. As a result, total gasoline tax revenue R will be, by definition, the tax rate per gallon T times total gasoline consumption. This relationship can be expressed as:

$$R \equiv \frac{TV}{e} \tag{1}$$

where V represents highway usage in vehicle miles and e, average gasoline efficiency measured in miles per gallon.

From the fact that gasoline tax revenues are dedicated by law to highway expenditures we obtain:

$$R = \mu H + k\frac{dH}{dt} \tag{2}$$

where H represents the size of the existing highway system; k, the average cost of construction per mile of new highway; and μ, the average maintenance cost per mile per year to prevent existing highway from deteriorating.[3]

---

* Associate professor, Carleton University, and associate professor, Graduate School of Business, University of Pittsburgh, respectively. With the usual *mea culpa* qualifications we would like to acknowledge the assistance of A. Reza, M. Spiro, and R. Bhaskar, of the University of Pittsburgh, S. B. Park of Carleton University, and the referee.

[1] Law now permits some use of the Federal Highway Trust Fund for mass transit. However, the amounts are rather small and the uses quite limited.

[2] William Shannon states: "Forty-five of the 50 states have either constitutional or statutory restrictions on the use of gasoline taxes for any purposes except highway construction and maintenance" (p. 120). Beginning with the Federal-Aid Highway Act of 1956, federal gasoline tax revenues were dedicated to financing highway construction.

[3] In practice, not all earmarked funds have been expended so that a surplus has been accumulated. The model could be adjusted to include a parameter multiplying R in equation (2) representing the fraction of current funds expended, and the analysis would not be

Highway usage is plausibly affected by a variety of factors in addition to the size of the highway system. General economic conditions affect the demand for vehicular transportation since transportation is both an input in production and a form of consumption. Recently, of course, the availability of gasoline has had a dramatic effect. Changes in the price of gasoline may also be important. In the long run, demographic and structural factors will play a role. For the purposes of this model, all factors affecting highway usage except the size of the highway system (and later, economic conditions) are considered exogenous.

The relationship between highway usage and highway size is expressed as:

$$(3) \quad V = h(H), \ h' > 0, \ h'' < 0, \ h(0) = 0$$

The assumptions about the derivatives of h can be rationalized by a diminishing returns argument. Although more roads generate greater usage, the most needed roads are built first.

Combining these three equations and assuming an initial value for H, the dynamic behavior of the model is described by:[4]

$$(4) \qquad \frac{dH}{dt} = \frac{1}{k}\left[\frac{Th(H)}{e} - \mu H\right]$$

The model has a unique, stable equilibrium,[5] given the assumptions of equation (3), if:

$$(5) \qquad \lim_{H \to 0} h'(H) > \frac{e\mu}{T}$$

If this inequality were not satisfied, even with an arbitrarily small highway network, revenues would not cover maintenance required to maintain the system.

The comparative static behavior of the model is easily seen. The higher the tax rate or the less efficient the vehicles, the greater the tax revenue and the larger the equilibrium highway system. The greater the maintenance costs, the smaller is the road network. Construction costs do not affect the equilibrium size; instead they determine the adjustment time to equilibrium.

II. Estimation

In this section an attempt is made to estimate the vehicle-usage function and to predict the size of the equilibrium highway system.

We postulate as a model for usage that:

$$(6) \qquad V = A - BH^{-\alpha} + CY$$

H enters in this exponential form because it is only reasonable that as H goes to infinity, usage should approach a finite limit.[6] Moreover, of all the other factors which may affect highway usage besides highway size, the state of the economy, measured by the level of real GNP (Y), would appear to be the most important variable in explaining year to year changes. With B and α constrained to be positive, this functional form is consistent with the assumptions of equation (3).

We wished to estimate this relationship for the period 1950 to 1969.[7] However, the time trend common to V, H, and Y makes direct estimation unwise. To minimize this problem first differences were used:

$$(7) \qquad V_t - V_{t-1} = -B[H_t^{-\alpha} - H_{t-1}^{-\alpha}] + C[Y_t - Y_{t-1}]$$

This equation was then estimated by a nonlinear least squares technique with the re-

---

substantively affected. It should be noted that in our model, new construction is a residual determined by the funds remaining after the necessary maintenance has been financed.

[4] The parameters k, T, e, and μ are taken to be constants. If these parameters behaved systematically with respect to time, there would in general be no finite, nonzero equilibrium.

[5] A precise characterization of behavior in the region to the right of the equilibrium would require an assumption about the rate at which highways crumble into nonexistence when they are not maintained.

[6] H would affect usage partly by reducing the money cost of a trip and partly by reducing the travel time per trip. Both types of costs would have a positive lower bound as H increases.

[7] If the years 1973 and 1974 were included in our estimation, then the availability and price of gasoline would also be of major significance in explaining usage. Bearing in mind these factors, one would want to reestimate our results before assigning much significance to our subsequent extrapolation for the equilibrium highway size.

sults as follows:[8,9]

$$(8)\quad V_t - V_{t-1} = -14{,}900[H_t^{-.04337} - H_{t-1}^{-.04337}] + 1.008[Y_t - Y_{t-1}], \quad \overline{R}^2 = .91$$

These estimates for B, α, C were then used in equation (6) and the difference between the actual and predicted values for V was taken as an estimate of the intercept A. The 19 estimates, so obtained, were then averaged to yield a point estimate for A of 11,945.[10]

It is possible to use this usage relationship to extrapolate to find an estimate for the equilibrium highway size (when revenues balance maintenance expenditures) if the parameters and Y are held constant. This involves solving for H_B from the equation:[11]

$$(9)\qquad A - BH_B^{-\alpha} + CY = \frac{e\mu}{T} H_B$$

This yields a value for H_B of 5.6 million miles, approximately double the current length.

III. Concluding Remarks

We have shown, with reference to the highway system, that use-financed construction programs imply a differential equation solution for capacity. Moreover, it is possible to estimate a usage relationship and obtain a tentative idea of future behavior by extrapolation.

One can generally criticize the earmarking of revenues from any tax as an unnecessary barrier to the efficient allocation of resources. The gasoline tax is no exception. However, assuming that some earmarking of tax revenues will continue, knowledge of the implicit time path of capacity would, at least, be useful to a policy maker with responsibility for choosing tax and expenditure parameters.

REFERENCES

H. Belloc, *The Highway and its Uses*, London 1926.

W. V. Shannon, "The Untrustworthy Highway Fund," *The New York Times Magazine*, Oct. 15, 1972.

U.S. Department of Commerce, *Survey of Current Business*, various issues.

U.S. Department of Transportation, Federal Highway Administration, *Highway Statistics*, summary to 1965 and individual volumes, 1966–69.

---

[8] Y was taken to be the level of real *GNP* in billions of dollars and V and H were represented by the series for vehicle miles (in billions per year) and miles of surfaced roads and streets (in tens of thousands), respectively.

[9] The program (Gaushaus from the University of Wisconsin Computing Center) prints out very approximate 95 percent confidence intervals based on a linear approximation in the neighborhood of the final parameter values. The intervals were $\pm 2.23 \times 10^6$, ± 1.205, and $\pm .293$, for B, α, and C, respectively. The large intervals for B and α are an unfortunate consequence of fitting a term of the form aX^b. When a linear regression was run with α set at .04337, the same coefficients were obtained (of course) for B and C and the corresponding t values were 3.60 and 7.19, respectively.

[10] This procedure seems appropriate so long as the random disturbance in equation (6) has zero expectation. The estimates for A ranged from a high of 11,989 to a low of 11,930.

[11] We used 1968 values for the parameters: T was \$.1062 per gallon; μ, \$2,065 per mile; e, 12.25 miles per gallon. Y_{1968} was 706.6 in billions of 1958 dollars.

Optimal Advertising: A Note

By DAVID WEBER*

In an article which appeared some twenty years ago, Robert Dorfman and Peter Steiner derived the necessary conditions for optimal advertising expenditures for the case of a firm which aims to maximize total profit and which is isolated from the competition of rivals. Their result called for that amount of advertising for which the marginal value product of advertising equals the absolute value of the partial price elasticity of demand. It may be shown that this result holds for other market circumstances as well (see my dissertation).

I should now like to demonstrate that an interesting variant of the Dorfman-Steiner condition holds if demand uncertainty is incorporated into an otherwise straightforward leader-follower model, as first developed by Stackelberg. I focus on the follower firm whose objective is assumed to be the maximization of expected total profit, to be defined below. In the interests of mathematical tractability, the demand function of the follower is assumed to be:

$$Q_2 = A^{\delta} P_2^{\alpha} P_1^{\sigma} a_2^{\beta} a_1^{\gamma} v \tag{1}$$

where v is a uniformly distributed random variable such that the probability density function

$$f(v) = \frac{1}{k} \qquad \text{if } 0 \leq v \leq k \tag{2}$$

$$f(v) = 0 \qquad \text{otherwise} \tag{3}$$

(I have chosen to use such an error, not only because manipulations are not difficult to perform, but, in addition, a uniform error is used in the model developed by Edwin Mills which I am following, and, in light of the general statement of my model, it is sufficient to permit the incorporation of demand uncertainty.) Also, Q_2 is the quantity demanded of the follower's product, P_2 is the price set by the follower, a_2 is the level of advertising set by the follower, P_1 is the price set by the leader, a_1 is the level of advertising set by the leader, and A measures real disposable income per capita. Following the framework developed by Mills, the follower is presumed to assume that for any price-output-advertising combination he might set, there is a chance he will be short by some amount, the mean amount of shortage for any such combination depending on the probability distribution of demand.

The follower's expected profit function is:

$$E(\pi_2) = E(P_2Q_2) - c_2z_2 - a_2 \tag{4}$$

where z_2 is the output of the follower, not likely to be the same as Q_2. That is, the follower's maximization problem depends on three independent decision variables, an optimal price-advertising combination no longer implying an optimal output. (Mills presumes that the output decision is made before the actual quantity demanded for the period, i.e., v, is revealed.) Now the first term on the right-hand side of (4) is expected total revenue, which Mills defines to be the difference between the value of expected demand and the value of expected shortage. Only if each possible quantity demanded (at a given price-advertising combination) could be sold would the value of mean demand and expected total revenue be the same; those quantities demanded that could not be satisfied from a given output should not be included in the follower's calculation of expected total revenue. Substituting the demand and density functions defined above into the follower's expected profit function, we have:

$$E(\pi_2) = A^{\delta} P_2^{\alpha+1} P_1^{\sigma} a_2^{\beta} a_1^{\gamma} \frac{k}{2} \tag{5}$$

(over)

---

* U.S. Coast Guard Academy. This note is based on my doctoral dissertation. I thank a referee for comments, but accept full responsibility for any errors that remain.

$$- P_2 \int_{z_2/f_2}^{k} \frac{f_2 v - z_2}{k}\, dv$$

$$- c_2 z_2 - a_2$$

where

$$(6) \qquad E(v) = \frac{k}{2}$$

$$(7) \qquad f_2 = A^{\delta} P_2^{\alpha} P_1^{\sigma} a_2^{\beta} a_1^{\gamma}$$

The value of the integral is the mean amount of shortage expected when P_2, a_2 is set and z_2 is produced. Shortage will occur when actual demand exceeds output, i.e., when $f_2 v > z_2$, or for those values of the random variable greater than z_2/f_2.

If we now take the partial derivative of (5) with respect to the follower's three decision variables and set each equal to zero, the resulting set of first-order conditions may be manipulated to give the desired result, a variant of the Dorfman-Steiner theorem for optimal advertising expenditure. This three-equation system may be written compactly as follows:

$$(8) \qquad f_2 \frac{k}{2} + P_2 \frac{k}{2} \frac{\partial f_2}{\partial P_2} = g_2 + P_2 \frac{\partial g_2}{\partial P_2}$$

$$(9) \qquad P_2 \left[1 - F\left(\frac{z_2}{f_2} \right) \right] = c_2$$

$$(10) \qquad P_2 \frac{\partial f_2}{\partial a_2} \left[\frac{k}{2} - \left[\frac{k^2 f_2^2 - z_2^2}{2kf_2^2} \right] \right] = 1$$

where $g_2 = \int_{z_2/f_2}^{k} \frac{f_2 v - z_2}{k}\, dv$

and $F(v)$ is the cumulative distribution function corresponding to $f(v)$. Multiplying the second term on the left-hand side of (8) by f_2/f_2 and substituting for $\partial g_2/\partial P_2$, (8) becomes:

$$(11) \quad f_2 \frac{k}{2} + h f_2 \frac{k}{2} = g_2 + P_2 \frac{\partial f_2}{\partial P_2} \left[\frac{k^2 f_2^2 - z_2^2}{2kf_2} \right]$$

where h is the partial price elasticity of demand of the follower's riskless demand curve. Multiplying the partial derivative on the right-hand side of (11) by f_2/f_2 and substituting from (10), (11) becomes:

$$(12) \quad f_2 \frac{k}{2} (h + 1) = g_2 + f_2 h \left[\frac{\mu \frac{k}{2} - 1}{\mu} \right]$$

where $\mu = P_2(\partial f_2/\partial a_2)$, which is what Dorfman and Steiner call the marginal value product of advertising. Equation (12) may be manipulated into:

$$(13) \qquad \frac{h}{\mu} = \frac{g_2}{f_2} - \frac{k}{2}$$

With uncertainty present, the relationship between h and μ depends on the price and output set. If we assume that P_2 and (or) z_2 is set so high that mean shortage g_2 is zero, the condition for optimal advertising is very similar to that derived by Dorfman and Steiner in which case certainty is assumed. (Note that $\mu k/2$ may be interpreted as the mean marginal value product of advertising, with $f_2 k/2$ being mean demand.) But if g_2 is not zero, which Mills argues is likely, h will be less than $\mu k/2$ in absolute terms at the optimal level of advertising since $g_2 \leq f_2 k/2$. (Note that $g_2 = f_2 k/2$ only if $z_2 = 0$.)

I conclude by demonstrating that associated with this result is the conclusion that with uncertainty present in the manner described, a smaller amount of advertising will be optimal. First we note that (10) may be interpreted as requiring that at the optimal level of advertising, the expected value of the marginal value product of advertising is greater than the marginal cost of an extra dollar of advertising, which is one, or that a fraction of the expected value of the marginal value product of advertising equals one at the optimal level of advertising. Second, it may be shown that if the assumption of uncertainty is dropped while all other features of the model remain intact, one of the two resulting first-order conditions requires that the marginal value product of advertising be one. If $\beta < 1$ per second-order conditions for either model, the marginal value product of advertising will be smaller only at a higher level of advertising. A rationale for less advertising being optimal under uncertainty is that because there is a positive probability of shortage, the follower is reluctant to advertise as much for fear that some of the demand which will be generated will not be able to be satisfied, given the output produced.

REFERENCES

R. Dorfman and P. Steiner, "Optimal Advertising and Optimal Quality," *Amer. Econ. Rev.*, Dec. 1954, *44*, 826–36.

E. Mills, *Price, Output, and Inventory Policy*, New York 1962.

D. Weber, "Advertising Behavior in Oligopolistic Markets," unpublished doctoral dissertation, Brown University 1973.

On the Effects of Fiscal and Monetary Policy: Completing the Taxonomy

By LAURENCE H. MEYER AND WILLIAM R. HART*

In a recent issue of this *Review*, Bent Hansen presented a taxonomic discussion of the effects of monetary and fiscal policies. Hansen suggests that the analysis "... bears directly upon the controversy between Keynesians and Monetarists" (p. 546). A central issue in this debate is the potency of bond financed fiscal operations and the nature and quantitative significance of the "crowding-out" effect. Complete crowding out occurs when an increase in government expenditures crowds out an equivalent amount of private expenditures, leaving aggregate demand and, hence, income unchanged. According to the Monetarists, this crowding out of private expenditures reflects the crowding out of private borrowing by the increased government borrowing. Hansen's model is designed to capture this interaction of government and private spending and borrowing and, on the surface, appears well suited for this purpose. It includes both government and private bond markets and imposes financing constraints on both government and private investment expenditures. Nevertheless, an increase in the supply of government bonds has no direct impact on the supply of or demand for private loans.

The unusual feature of Hansen's model which prevents it from capturing this effect is the lack of a direct connection between the two bond markets. Because all private investment is financed through bank loans and because banks do not hold government bonds, Hansen's model effectively dichotomizes the financing of private investment and the financing of government expenditures. As a result, the government does not compete directly with the business sector for sources of funds in the financing of expenditures.

In this comment we will modify Hansen's model so that it can capture the direct portfolio impact of an increase in the supply of government bonds on the private loan rate. First we will present Hansen's model and consider its relation both to the Monetarist view of the crowding-out effect and to recent attempts to model this effect.

I. Hansen's Model

The general form of the model used by Hansen in discussing the taxonomy is shown below.

$$(1)\qquad Y = I + C + G$$

$$(2)\qquad C = c_1(Y - T + B) + c_2(M^n + \alpha M^b + B/r_g) + c_3; \quad c_1 > 0, \quad c_2 > 0$$

$$(3)\qquad I = i_1 r_b + i_2 Y + i_3; \quad i_1 < 0,\ i_2 > 0$$

$$(4)\qquad T = t_1 Y + t_2; \qquad t_1 > 0$$

$$(5)\qquad M^n = nY; \qquad n > 0$$

$$(6)\qquad \alpha M^b = m_1(r_g - \mu r_b) + m_2 Y + m_3 B/r_g; \quad m_1 < 0, \quad m_2 > 0, \quad m_3 > 0$$

$$(7)\qquad \bar{K} + I = (\alpha - 1)M^b + L^b$$

$$(8)\qquad L^b = \beta_1(r_b - \bar{r}) + \beta_2; \qquad \beta_1 > 0$$

All variables in the model are as defined by Hansen: Y is income, I investment, C consumption, G government expenditures, T tax revenue, B is the number of government bonds held by the private sector (perpetuities with \$1.00 interest payments per period), M^n currency in circulation, M^b commercial bank deposits at the central bank, r_g bond market rate of interest, r_b commercial banks' loan rate, α inverse of the required reserve ratio, μ the ratio between commercial banks' deposit and loan rates, $\overline{K}$ initial private stock of capital, L^b rediscounts and advances to commercial banks, $\bar{r}$ the central bank's discount rate, D_0 the budget deficit at the beginning of the period.

---

* Assistant professor, department of economics, Washington University, St. Louis, and instructor, department of economics, Miami University, Ohio, respectively. We would like to thank Bent Hansen for providing us with his worksheets detailing the computation of his multipliers. We also received helpful comments from the participants in the Research Workshop in Money and Banking at Washington University. An appendix containing the computation of the multipliers discussed in this paper is available upon request.

The model is completed with the addition of a public sector budget constraint. Two possible public sector budget constraints are considered by Hansen. The European constraint reflects an institutional setting in which the central bank issues currency, makes loans to and keeps deposits for commercial banks, and acts as a banker for the government. This constraint takes the following form:

$$\text{(9)} \quad dM^n + dM^b = dG - dT + dB + D_0 - dB/r_g + dL^b$$

The *U.S.* constraint reflects an institutional setting in which the government does not use the central bank and therefore is forced to finance its deficits and surpluses exclusively through bond transactions with the private sector. This constraint takes the following form:

$$\text{(9a)} \quad dM^n + dM^b = dL^b - dB_{CB}/r_g$$

$$\text{(9b)} \quad dB_{Tr}/r_g = dG - dT + dB + D_0$$

$$\text{(9c)} \quad dB = dB_{CB} + dB_{Tr}$$

where $-dB_{CB}$ is the number of government bonds bought by the central bank in the open market and dB_{Tr} is the number of bonds sold by the Treasury to the private sector.

Although the public sector budget constraint is not new to economic analysis, Hansen's discussion is unique because he points out that some knowledge of the institutional set-up of the economy is required to determine the precise specification of the constraint. The specific form of the budget constraint, in turn, builds into the model the limitations in policy options imposed by institutional arrangements. A second unique and innovative feature of Hansen's model is the requirement that all investment be financed through bank loans. This requirement is embodied in equation (7), the equilibrium condition in the bank loan market. Hansen's model thus extends the logic of the public sector budget constraint to the private sector.

Equations (1)–(8) and (9) represent what Hansen calls the European model, while equations (1)–(8) and (9a), (9b), (9c) represent the *U.S.* model. Our principal concern is with Hansen's failure to incorporate the role of credit markets in the crowding-out effect. Therefore we concentrate on the effects of a bond financed increase in government expenditures. Using the European version of the model, G, t_1, B, $\bar{r}$, and α are selected as policy instruments, and their corresponding differentials represent the discretionary measures taken. To determine the impact of a bond financed fiscal operation, we let all discretionary policies and autonomous changes, other than dG and dB, equal zero. Solving the system for dY, the bond financed operation yields:

$$\text{(10)} \quad dY = \frac{1}{\Delta} \cdot \Big\{ \underset{I}{dG} + \underset{II}{c_1 dB + c_{\_}dB/r_g} + \underset{III}{dB/r_g(\delta_2)} \Big\}$$

where $1/\Delta =$ the basic multiplier and is assumed positive and

$$\delta_2 = \frac{m_3 c_2 (B/r_g^2)}{m_1 - m_3(B/r_g^2)} < 0$$

The multiplicand consists of three terms, each one measuring an impact on output. Term *I* is the direct impact on demand by virtue of the fact that G is a component of aggregate demand. Terms *II* and *III* represent two impacts resulting from the increase in the supply of government bonds. The total impact on income is ambiguous.

In light of the arguments by Monetarists and recent work concerning the financing of government expenditures (see articles by William Silber and Meyer), this does not appear to be too startling a result, and one might be led to believe that terms *II* and *III* are, in fact, capturing the impact of credit markets in the crowding-out effect. This result does appear quite startling, however, if we let c_2, the marginal propensity to consume out of wealth, and the marginal propensity to consume out of interest payments[1] go to zero. The solution dY becomes:

$$dY = \frac{1}{\Delta} \cdot \{dG\} \tag{10a}$$

which is unambiguously positive. This implies that in the absence of a wealth effect in the consumption function there is no direct impact on income resulting from the financing of government expenditures. This result certainly contradicts both recent Monetarist discussions (see, for example, Karl Brunner and Allan Meltzer, p. 953; Milton Friedman, p. 51; and John Culbertson, p. 463) of and recent attempts to model (see articles by Silber and Meyer) the crowding-out effect in which crowding out occurs independently of any wealth effect in the consumption function. While Hansen has offered an additional channel through which crowding out can occur, namely through the wealth effect in the consumption function, he has in the process omitted the channel most frequently emphasized by Monetarists: the direct impact of an increase in the supply of government bonds on interest rates in private credit markets.

II. The Crowding Out of Private Borrowers

A major argument of the Monetarists is the failure of income expenditure models to explicitly consider the impact on credit markets resulting from the financing of a government deficit. The design of Hansen's model is intended to capture the impact of the financing of government deficits on the private credit market. Its specification of portfolio behavior, however, prevents it from doing so.

A bond financed fiscal operation can be described in terms of direct impacts and feedback responses.[2] Most macro models include a single direct effect: the *direct fiscal impact*, (*D1*)—the direct impact of an increase in G on aggregate demand by virtue of the fact that G is a component of aggregate demand. In such models, a bond financed fiscal policy is unambiguously expansionary unless one or more parameters take on extreme values (for example, zero interest elasticity of demand for money or infinite interest elasticity of investment). The only way crowding out can occur in such models is through the negative feedback response. As this response is activated by an increase in income, it can at best prevent income from changing; it can never generate a decline in income.

Recent contributions by Silber and Meyer have introduced two additional direct effects, both arising from the increase in the supply of government bonds: the *direct wealth impact*, (*D2*) and the *direct portfolio impact*, (*D3*). These effects were ignored in previous models either because of the failure to take into account the government budget constraint or because of specifications of the consumption and money demand functions that excluded any direct influence of changes in the government bond supply. The direct wealth impact is the increase in consumption induced by the increase in wealth associated with the increase in the supply of government bonds. The direct portfolio impact arises from the direct impact of an increase in government bonds on interest rates. An increase in the supply of government bonds generates excess supply in the bond market, driving up the interest rate. The increase in the interest rate reduces investment. The direct wealth impact is expansionary but the direct portfolio impact is restrictive. Therefore, the net effect of a bond financed fiscal operation is ambiguous. Complete crowding out can occur without any parameters taking on extreme values and a bond financed increase in G may even result in a decline in income.

---

[1] An increase in the supply of government bonds increases consumption through the increase in disposable income (coupon payments) and through the increase in wealth (discounted value of current and future coupon payments). This appears to be double counting. To avoid this problem we eliminate the c_1dB terms in Hansen's multipliers.

[2] Direct impacts refer to the *initial* impact of the increase in government expenditures and increase in the supply of government bonds on aggregate demand and hence income. These initial impacts can be on consumption via changes in disposable income and/or wealth associated with the operation, on investment via changes in interest rates directly associated with the operation, or on government expenditures. Feedback responses arise in response to the initial changes in income. There are two major feedback responses: a positive feedback response on components of aggregate demand which are positive functions of income, and negative feedback effects due to the income induced rise in interest rates on components of aggregate demand that are negatively related to interest rates. In Hansen's multipliers, the parameters that govern the magnitude of the direct impacts appear in the multiplicand, and the parameters that govern the feedback responses appear in the basic multiplier, $1/\Delta$.

Hansen's model includes the direct fiscal and direct wealth impacts but excludes the direct portfolio impact on investment. Why then does he also get an ambiguous multiplier and how can his model, with its detailed treatment of both government and private financing of expenditures, fail to capture the direct impact of an increase in government bonds on the private bond rate? The ambiguous sign of his multiplier results from a fourth direct impact, excluded in previous models: *the direct interest-induced wealth impact on consumption,* ($D4$)—the impact on consumption of the decline in wealth due to the direct impact of the increase in government bond supply on the government bond rate. This effect is restrictive.

This is not, however, the channel emphasized by Monetarists. For example, note the mechanism emphasized by Culbertson:

> Fiscal policy provides additional spending in a world of sparse spending opportunities. But it does not provide a new source of finance in a world where spending is constrained by sources of finance. The government expenditures are financed in debt markets in competition with private expenditures. The case least favorable to fiscal policy is that in which the additional government borrowing simply crowds out an equal (or conceivably greater) volume of borrowing that would have financed private expenditures. [p. 463]

Hansen apparently believes he has captured the mechanism referred to by Culbertson: "Our model thus allows for the 'crowding-out' from the financial markets of private borrowers when the government finances a deficit by sales of securities" (p. 552, fn. 10). Yet if $c_2 = 0$ in Hansen's model,[3] the effect on income of a bond financed increase in G is unambiguously positive, in contrast with the direct quote above by Culbertson and the models of Silber and Meyer.[4]

Hansen's model fails to generate the direct portfolio impact on investment because the nonbank private sector holds money and government bonds and the banking system holds reserves and private bonds (loans), but no one holds both government and private bonds. An increase in the supply of bonds increases the government bond rate but does not put upward pressure on the private bond (loan) rate. Despite what appears to be the attractive feature of including both government and private bond markets, the failure to thoroughly interconnect the two markets prevents the model from capturing the crucial role of the credit market in the crowding-out effect. This defect, however, can be easily remedied.

III. Respecification of Hansen's Model

If we respecify the model to allow both households and commercial banks to hold government bonds, the government is forced to compete directly with the private sector for sources of finance. This competition, in turn, leads to the type of financial market crowding out the Monetarists have emphasized.

This respecification necessitates the following changes in Hansen's basic model (his equations (1)–(8) plus (9′), p. 553): Hansen's equation (6) is replaced by:

---

[3] Whenever we set $c_2 = 0$ we assume the marginal propensity to consume out of interest payments is also zero. We acknowledge that Hansen has captured all the wealth effects generated by a bond financed fiscal policy. We set $c_2 = 0$ so we can concentrate on other effects he has omitted.

[4] Note further that if we set $c_2 = 0$ the multiplier for the financing of an initial deficit D_0 by issuing new bonds is

$$dY = \{(D_0 + dB)(\delta_3) - dB/r_g(\delta_3)\}$$

where δ_3 is positive. The increase in the supply of bonds must equal the coupon payments on the new bonds (dB) as well as the initial deficit. Therefore $dB/r_g = D_0 + dB$ and $dY = 0$.

$$(6') \quad \alpha M^b = m_1(r_g - \mu r_b) + m_2 Y + m_3 Bp/r_g$$

where Bp = the number of government bonds held by the household sector.

Hansen's equation (7) is replaced by:

$$(7') \quad \bar{K} + I = (\alpha - 1)M^b - Bc/r + L^b$$

where Bc = the number of government bonds held by commercial banks. This relation still assumes that all private investment is financed through bank loans, but as commercial banks are now allowed to hold government securities any bonds held by commercial banks reduce the availability of funds to the business sector.

Two new equations are required to complete the model:

$$(11) \quad Bc/r_g = e_1(r_g - r_b) + e_2$$

$$(12) \quad B = Bc + Bp$$

Equation (11) specifies the commercial banks' demand for government bonds as a positive function of the difference between the government bond and bank loan rates. Equation (12) requires that the holdings of government bonds by households and commercial banks sum to the total stock outstanding.

Finally, we assume there is no wealth effect in the consumption function and equation (2) is replaced by:

$$(2') \quad C = c_1(Y - T) + c_3$$

Letting all policy instruments, other than dG and dB, and all autonomous changes go to zero, we solve for the effect of a bond financed increase in government expenditures on income:

$$(13) \quad dY = \frac{1}{\Delta''}\{dG - [dB/r_g](\pi_3)\}$$

The basic multiplier $1/\Delta''$ is assumed to be positive and

$$\pi_3 = i_1 e_1 m_3 \div [m_4(\alpha\beta_1 - i_1 + e_1) + e_1(m_3 i_1 - \alpha\beta_1 m_3 - m_1\mu - \alpha\beta_1)] > 0$$

assuming $1 > \mu > 0$; i.e., banks' loan rate exceeds their deposit rate.

The two terms in the bracketed expression of equation (13) measure the impact of the bond financed operation. The first term, dG, is the direct fiscal impact; it is positive. The second term, $-(dB/r_g)(\pi_3)$, captures the direct portfolio impact on investment of the bond financing; it is negative. The increase in the supply of government bonds now puts upward pressure on both r_g and r_b which in turn puts downward pressure on, and crowds out, private investment. This crowding out of private investment is captured by the coefficient i_1 in π_3.[5]

There are, of course, other ways the model could have been specified so as to yield similar results. The portfolio behavior of firms could be altered by letting them, rather than commercial banks, hold government bonds in their portfolios. The effect of this modification would be to introduce the government bond rate, a measure of the opportunity cost of purchasing additional capital, as an argument in the investment function. Any change in the government bond rate would then have a direct impact on investment. Alternatively, we could have assumed firms finance all investment expenditures exclusively through the issue of private bonds to households. Households who now hold private as well as government bonds would provide the competitive link between the financing of private investment and the financing of government expenditures.[6]

---

[5] Note, further, that the multiplier for the financing of an initial deficit by issuing new bonds is now

$$dY = \frac{1}{\Delta''}\{-dB/r_g(\pi_3)\}$$

which is unambiguously negative under the assumption that $c_2 = 0$. Recall that this expression for dY would be zero in Hansen's model if $c_2 = 0$.

[6] It is interesting to note, however, that in the latter case the direct portfolio impact on investment is no longer unambiguously restrictive. The increase in the supply of government bonds will have two effects: 1) there will be a reduction in the demand for private bonds due to the increase in the government bond rate (substitution effect), but 2) there will be an increase in the demand for private bonds due to the increase in wealth associated with the increase in the supply of government bonds (wealth effect). The former impact puts upward pressure on the private bond rate while the latter impact puts downward pressure on this rate. The net impact on the private bond rate, and hence investment, is ambiguous.

IV. Bond Financed Fiscal Policy in the *U.S.* Model

Hansen points out that "It is immediately clear that what can be done under one of these budget constraints can also be done under the other . . . " (p. 550). A bond financed increase in government expenditures should yield identical results under the two regimes. However, the multiplier for such an operation while ambiguous in the European model (see equation (10) above) is unambiguously positive in the *U.S.* model! The multiplier in the *U.S.* model is

$$dY = \frac{1}{\Delta}\{(1 + \delta_3')dG\} \tag{14}$$

Hansen claims that δ_3' is unambiguously positive so that $dY/dG > 0$. This result, if correct, implies that the specific form of the budget constraint might be an important issue in the controversy between Keynesians and Monetarists. The contradiction, however, is erroneous, arising from Hansen's misreading the sign of the δ_3' term in (14). The full expression for δ_3' is:

$$\delta_3' = (c_2 B/r_g^2)\frac{m_3}{m_4(1 - r_g)} + \frac{c_1 r_g + c_2}{1 - r_g} \tag{15}$$

The first term on the right is negative ($c_2 > 0$, $m_3 > 0$, but $m_4 < 0$) and the second term is positive. The δ_3' term includes the two wealth effects (and the additional impact via the coupon payments) associated with an increase in the supply of government bonds: the first term is the restrictive impact of a higher government bond rate on consumption via the interest-induced wealth effect, and the second term is the positive wealth effect due to the increase in the number of government bonds at the initial interest rate. The multiplier, of course, still is missing the direct portfolio impact on investment.

V. Conclusion

Hansen's paper includes two interesting innovations: the requirement that all investment be financed through bank loans and the limitation on policy options imposed by institutional arrangements. Hansen fails to exploit the former and reports incorrect results concerning the implications of the latter. Readers may, therefore, have failed to appreciate the value of Hansen's innovations. We hope our comments help to exploit Hansen's innovations and contribute toward completing the taxonomy he began.

REFERENCES

K. Brunner and A. Meltzer, "Money, Debt, and Economic Activity," *J. Polit. Econ.*, Sept./Oct. 1972, *80*, 951–77.

J. Culbertson, *Macroeconomic Theory and Stabilization Policy*, New York 1968.

M. Friedman, "Has Fiscal Policy Been Oversold?" in M. Friedman and W. Heller, eds., *Monetary vs. Fiscal Policy*, New York 1969.

B. Hansen, "On the Effects of Fiscal and Monetary Policy: A Taxonomic Discussion," *Amer. Econ. Rev.*, Sept. 1973, *63*, 546–71.

L. Meyer, "The Balance Sheet Identity, The Government Financing Constraint, and The Crowding-Out Effect," *J. Monetary Econ.*, Jan. 1975, *1*, 65–78.

W. Silber, "Fiscal Policy in IS-LM Analysis: A Correction," *J. Money, Credit, Banking*, Nov. 1970, *2*, 461–72.

Default Risk, Scale, and the Homemade Leverage Theorem: Note

By PAO LUN CHENG*

In a recent article in this *Review*, Vernon Smith presented three theorems of considerable significance concerning scale, leverage, default risk, and expected utility of investor's final wealth contingencies. The theorems are derived from a model in which there is a single corporation[1] and a number of investors not necessarily having identical utility functions of wealth. The corporation makes all financial decisions at the beginning of a single interval of real investment. At the end of the period, in accordance with the contractual claim conditions of securities, it distributes a stochastic cash flow X among investors not necessarily having identical subjective distributions of X. Smith restricted himself to dealing with one class of debt whose claim on X is strictly prior to that of a single class of common shares. In addition to these two assets, each investor can hold a third asset, cash, with zero certain return. He showed that an investor's expected utility depends on the financial aggregates Y, Z, and M, where Y and Z are respectively the equilibrium market value of shares and bonds and $M=Y+Z$ is the scale of investment by the firm to generate X. Smith emphasized the fact that Y, Z, M, and $\mu=Z/Y$, the market leverage, represent externalities to the firm as well as to the investors. His theorems have to do with the effect of a change in the market leverage on an investor's expected utility.

With respect to Theorems 1 and 2, Smith disallowed the existence of a loan market for personal borrowing and lending, as compared with the models, for example, of Franco Modigliani and Merton Miller, Jan Mossin, John Lintner, and others, who assumed unrationed borrowing by investors at a riskless rate. For Theorem 3, however, Smith permitted the investors to borrow on personal account to buy shares without rationing at a rate which may be greater than, equal to, or less than the nominal lending rate on corporate bonds. Finally, for Theorems 2 and 3, the assumption of stochastic constant returns to scale is made, i.e., the investors' expected utilities are independent of M.

In an earlier paper, Smith (1970) expounded a Fundamental Leverage Theorem, which is essentially his Theorem 2. A comment by Uri Ben-Zion and Michael Balch on the theorem showed that it cannot hold if a change in the market leverage μ affects the nominal corporate bond rate $R-1$. Unfortunately, Ben-Zion and Balch, p. 293, fn. 3, were unable to provide the crucial proof of the intuitively appealing proposition that $dR/d\mu>0$ under the condition of positive default risk. The purpose of this comment is to show that $dR/d\mu>0$ when $Pr(X<X^*)>0$, where X^* is the firm's minimum cash flow required to escape default. If $dR/d\mu>0$, Smith's Theorems 2 and 3 will not hold, but his Theorem 1 stands, since the latter theorem assumes no default risk, i.e., $dR/d\mu=0$. In order to demonstrate the importance of the sign of $dR/d\mu$ to his theorems, we begin with his equation (noted as (6) in his text), the expected utility of an investor's final wealth.

$$(1)\qquad V=\int_{-\infty}^{0}U(x)H(M,X)dX+\int_{0}^{X^*}U(W')H(M,X)dX+\int_{X^*}^{\infty}U(W'')H(M,X)dX$$

---

* Professor, department of economics and commerce, Simon Fraser University.

[1] Later in the article Smith generalized Theorem 2 for the case of two corporations.

where

$H(M, X)$ = the investor's subjective probability distribution of X, associated with a scale M
$W' = X(z/Z)+x$
$= [(1+\mu)zX/\mu M]+x$, his final wealth if $0<X\leq ZR$
$W'' = zR+(X-ZR)(y/Y)+x$
$= zR+[1+\mu)yX/M]-\mu Ry+x$, his final wealth if $ZR<X$
Z, Y = market values of bonds and shares, respectively
x, y, z = his cash holdings, market value of share holdings, and market value of bond holdings, respectively
$X^* = ZR$, the firm's contractual obligations to bondholders
$R = X^*/Z$, one plus the contractual or nominal rate of interest on debt Z

With the way V, W', and W'' are specified, it is obvious both R and X^* affect V as well. Moreover, although Smith cogently argued that Z and Y are externalities, he seemed to assume implicitly that the nominal rate $R-1$ can be selected by the firm such that the market value of bonds Z will equal the face value of bonds, with the result that ZR represents the total payments to bondholders if the firm escapes default. But Z is an externality and one cannot expect the firm to select R such that $ZR=X^*$. Thus, a reasonable way to interpret Smith's specifications in this regard is to treat X^* as the total face value of bond certificates bearing *no* coupon interest, or equivalently, X^* is the number of discount bonds each having a face value of one dollar, entitling the holder of one bond to one dollar at the end of the period if the firm escapes default. We can then let Z be the market value of X^* to be determined as an externality and treat $R=X^*/Z$ as one plus the nominal market bond yield for the period. It is to be recognized that R represents a function of the internal decision variable X^* of the firm as well as the external market variable Z. We shall show that Z, Y, μ, and R depend on X^* and $dR/d\mu>0$ if the default risk is positive, after which the effect of $dR/d\mu>0$ on Smith's theorems will then be discussed in full.

Let the random variables Q and G denote the total return to bondholders and shareholders, respectively:

$$(2)\qquad Q = \begin{cases} 0, & X \leq 0 \\ X, & 0 < X \leq X^* \\ X^*, & X^* < X \end{cases}$$

$$(3)\qquad G = \begin{cases} 0, & X \leq X^* \\ X - X^*, & X^* < X \end{cases}$$

Although the investors may differ with respect to their subjective distributions of X, and hence, Q and G, they must obviously all agree that $Q+G=X$. Note that Q and G are conditional upon the choice of X^* by the corporation.

Suppose, instead of X^*, the firm chooses to issue X' discount bonds, $X'>X^*$. We have, instead:

$$(4)\qquad Q' = \begin{cases} 0, & X \leq 0 \\ X, & 0 < X \leq X' \\ X', & X' < X \end{cases}$$

$$(5)\qquad G' = \begin{cases} 0, & X \leq X' \\ X - X', & X' < X \end{cases}$$

As before, $Q'+G'=X$. Since there is no reason to expect a change in the investor's subjective distributions of X due to the change of X^* to X', each investor will notice that $F(Q')\leq F(Q)$ and $F(G')\geq F(G)$, where $F(\cdot)$ represent distribution functions. That is, for each investor, the condition of first-degree stochastic dominance exists[2] so that Q' is preferred to Q and G preferred to G'. The change in absolute preference in the same direction for every investor must imply that the equilibrium market values of Q' and G', say Z' and Y', must be such that $Z'/Y'>Z/Y$. Hence $\mu'>\mu$, implying $d\mu/dX^*>0$.

To reinforce our argument, in lieu of issuing a single class of X' discount bonds, the firm can choose to issue two classes of discount bonds plus shares: a first class of X^* discount bonds whose return is represented by the random variable Q, another class of $X'-X^*$ subordinate discount bonds whose return is represented by Q_1 where

---

[2] See Josef Hadar and William Russell, p. 27.

$$(6) \quad Q_1 = \begin{cases} 0, & X \leq 0 \\ 0, & 0 \leq X < X^* \\ X - X^*, & X^* \leq X < X' \\ X' - X^*, & X' \leq X \end{cases}$$

It is important to note that by referring to (2)–(6), $Q_1+G'=G$, $Q_1+Q=Q'$, and $Q+Q_1+G'=Q+G=Q'+G'=X$. In other words, if the firm chooses to issue a single class of X^* discount bonds, the return to shares G can be separated into Q_1 and G'. For example, suppose an investor i owns a fraction z_i/Z of X^* and a fraction y_i/Y of old shares with return G. The firm can "call in" all the old shares and replace them with $X'-X^*$ subordinate bonds with return Q_1 and new shares with return G', and remit to each investor i a fraction y_i/Y of subordinate bonds and new shares. Obviously, the investor's wealth contigency is unchanged, for $(y_i/Y)(Q_1+G')=(y_i/Y)G$. Note that the firm has succeeded in increasing its leverage, for $\Sigma_i(y_i/Y)\cdot(Q_1+G')+Q=Q_1+G'+Q=Q'+G'$. That is, the issuance of two classes of bonds is equivalent to the issuance of a single class of X' discount bonds.[3] Thus, with altogether three classes of securities with returns represented by Q, Q_1, and G', the equivalent of any two-class bonds and shares financing composition can be created. In the market, since $Q+Q_1=Q'$, the total value of X^* discount bonds and $X'-X^*$ subordinate bonds $Z+Z_1$ must be equal to the value of X' single class of bonds Z', i.e., $Z+Z_1=Z'$. By the same token, since $Q_1+G'=G$, the market value of G' must be $Y'=Y-(Z'-Z)=M-Z'$. It follows that

$$\mu' = \frac{Z'}{Y'} = \frac{Z'}{M - Z'} = \frac{Z + Z_1}{M - Z - Z_1}$$
$$= \frac{Z + Z_1}{Y - Z_1} > \mu = \frac{z}{y}$$

which is equivalent to $d\mu/dX^*>0$. Moreover, since $Z'=Z+Z_1$ so that $Z'>Z$, we can infer that $dZ/dX^*>0$. Similarly, since $Y=Y'+(Z'-Z)$ so that $Y>Y'$, whence $dY/dX'<0$.

In order to show $dR/d\mu>0$ when the default risk is positive, we need only to show $dR/dX^*>0$, since $d\mu/dX^*>0$ has already been established. Consider the case of X^* discount bonds and $X'-X^*$ subordinate bonds whose returns are represented by Q and Q_1, respectively. Let Q/X^* and $Q_1/(X'-X^*)$ be their respective returns per bond. Note that, using (2) and (6),

$$Q/X^* = \begin{cases} 0, & X < 0 \\ X/X^*, & 0 \leq X < X^* \\ 1, & X^* \leq X < X' \\ 1, & X' \leq X \end{cases}$$

$$Q_1/(X' - X^*) = \begin{cases} 0, & X < 0 \\ 0, & 0 \leq X < X^* \\ \dfrac{X - X^*}{X' - X^*}, & X^* \leq X < X' \\ 1, & X' \leq X \end{cases}$$

Again, $F(Q/X^*) \leq F(Q_1/(X'-X^*))$ so that Q/X^* stochastically dominates $Q_1/(X'-X^*)$. Hence, each investor absolutely prefers Q/X^* to $Q_1/(X'-X^*)$ so that the prices of bonds will be $Z/X^*>Z_1/(X'-X^*)$. Equivalently, we have

$$(7) \quad \frac{X^*}{Z} < \frac{X' - X^*}{Z_1}$$

suggesting a greater nominal rate of return for the subordinate bonds. Since the choice of issuing two classes of bonds with random returns Q and Q_1 is equivalent to the choice of issuing a single class of X' bonds with return Q', and $Z+Z_1=Z'$, the nominal return of Q' is

$$\frac{X'}{Z'} = \frac{X^* + (X' - X^*)}{Z + Z_1}$$

In conjunction with (7), we get

$$(8) \quad \frac{X^*}{Z} < \frac{X^* + (X' - X^*)}{Z + Z_1} < \frac{X' - X^*}{Z_1}$$

---

[3] If an equilibrium already exists before the new bond distribution, then the same equilibrium must exist after the new bond distribution. The existence of the prior equilibrium is implied by the comparative static analyses by Smith throughout his article.

Thus, the bond rate of a single class of X' discount bonds is greater than the rate of a single class of X^* discount bonds but is less than the rate of the subordinate bonds. It follows that $dR/dX^*>0$.

With $dR/d\mu>0$ established when the default risk is positive, we return to Smith's expected utility function (1) for an examination of his theorems. The investor selects (x^0, y^0, z^0) so as to maximize V subject to his initial wealth $W_0=x+y+z$, giving the Lagrangian $L=V+\lambda(W_0-x-y-z)$. The firm issues a single class of X^* discount bonds, plus shares. As before, Z and Y are the equilibrium market values of Q and G, respectively. Following Smith (p. 69), the Kuhn-Tucker conditions are

$$(9) \quad L_{x^0} \leq 0, \qquad x^0L_{x^0} = 0$$

$$(10) \quad L_{y^0} = I_1 + \mu I_2 - \lambda \leq 0, \qquad y^0L_{y^0} = 0$$

$$(11) \quad L_{z^0} = [(1+\mu)J_0/\mu] + J_1 - \lambda \leq 0, \qquad z^0L_{z^0} = 0$$

where

$$(12) \quad I_1 = \int_{X^*}^{\infty} U'(W'')(X/M)HdX$$

$$I_2 = \int_{X^*}^{\infty} U'(W'')[(X/M) - R]HdX$$

$$J_0 = \int_{0}^{X^*} U'(W')(X/M)HdX$$

$$J_1 = \int_{X^*}^{\infty} U'(W'')RHdX$$

Let us write the envelope equation

$$(13) \quad \frac{dV^0}{d\mu} = V^0_\mu + V^0_M\left(\frac{dM}{d\mu}\right) + V^0_R\left(\frac{dR}{d\mu}\right)$$

Smith's Theorem 1 states (p. 70):

> If there is no default risk on corporate bonds, an investor's optimal portfolio will be such that he will never prefer a decrease in the debt-equity ratio, and if $z^0>0$ he will be indifferent to a change in the debt-equity ratio, i.e.,

$$dV^0/d\mu = V^0_\mu + V^0_M(dM/d\mu) = 0$$

Having successfully demonstrated (p. 70) that $dM/d\mu=0$ when there is no default risk and since $dR/d\mu=0$, which is implicitly and correctly assumed in this case, (13) reduces to

$$dV^0/d\mu = V^0_\mu = -(z^0/\mu^2)J_0 + y^0I_2$$

By substituting from (9)–(12), he obtains

$$(1+\mu)V^0_\mu = -y^0L_{z^0} \geq 0, \qquad z^0L_{z^0} = 0$$

and Theorem 1 follows.

Let us now use (13) to examine his Theorem 2 (p. 71):

> If a corporation can invest at constant return to scale, and the default risk on its bonds is positive, then an investor's optimal portfolio will have the property that he will prefer the corporation to increase, leave unchanged or decrease its debt-equity ratio according as $\mu \gtreqless (z^0/y^0)$.

Since $V^0_M=0$ follows from the assumption of constant return to scale while retaining the implicit but incorrect assumption $dR/d\mu=0$ in this case of positive default risk, he continued to work with $dV^0/d\mu=V^0_M$. Substituting from (9)–(12), he obtained

$$dV^0/d\mu = V^0_\mu = (\lambda - J_1)(y^0 - z^0/\mu)/(1+\mu)$$

Using the inequality in (11) $\lambda-J_1\geq(1+\mu)J_0/\mu$, he reached the conclusion in Theorem 2. However evaluating (13) by dV^0/dX^* and then dividing the result by $d\mu/dX^*>0$, we obtain instead

$$(14) \quad dV^0/d\mu = (\lambda - J_1)(y^0 - z^0/\mu)/(1+\mu) + (z^0 - \mu y^0)\cdot\int_{X^*}^{\infty}(dR/d\mu)U'(W'')HdX$$

Note that the second term will have a sign always opposite to the sign of the first term, for as $(y^0-z^0/\mu)\gtreqless 0$, $(z^0-\mu y^0)\lesseqgtr 0$, while $dR/d\mu>0$ and $U'(W'')>0$. Consequently, Theorem 2 does not hold.

It is interesting to rewrite (14) as

$$(15) \quad dV^0/d\mu = (y^0 - z^0/\mu)\int_{X^*}^{\infty}\{(X/M) - R[1+(dR/R)/(d\mu/\mu)]\}U'(W'')HdX$$

using the fact that

$$(\lambda - J_1)/(1+\mu) = I_2 = \int_{X^*}^{\infty} U'(W'')[(X/M) - R]H\,dX$$

when $L_y^0=0$, $y^0>0$ so that $\lambda=J_1+(1+\mu)I_2$. It is clear that the sign of $dV^0/d\mu$ depends not only on the sign of (y^0-z^0/μ), but also on the magnitude of the "Leverage elasticity" of the nominal bond yield $(dR/R)/(d\mu/\mu)$. For example, if the elasticity is zero under the condition of zero default risk, (15) reduces to

$$\begin{aligned} dV^0/d\mu &= (y^0 - z^0/\mu)I_2 \\ &= -(z^0/\mu^2)J_0 + y^0 I_2 \\ &= y^0 I_2 \end{aligned}$$

where $J_0=0$ when bonds are riskless. Note that this is precisely the case for Smith's Theorem 1. However, when the elasticity is positive, or equivalently, $dR/d\mu>0$, the integral in (15) can be of any sign, depending on the choice of X^*. Therefore, we cannot a priori assert the direction of an investor's preference for leverage solely by comparing μ and z^0/y^0 as Theorem 2 would have us believe.

Let us now turn to Smith's Theorem 3 dealing with investors who are allowed to borrow to buy shares. Having borrowed an additional amount x to buy shares at a contractual rate $R'-1$, an investor after paying off his debt and interest will find his final wealth exactly zero if $(X-X^*)y/Y=xR'$. In other words, if $X=X_0\equiv(xR'+\mu Ry)Y/y$ the investor's final wealth is zero. Defining $\theta=(X/M)-1$, we see that the critical value of θ for the investor, corresponding to X_0, is

$$\theta' = [1/(1+\mu)][(x/y)R' + \mu R] - 1$$

Following from the definition of θ and θ', Smith characterized the investor's final wealth as follows:

$$W' = 0, \qquad \text{if } \theta \leq \theta'$$

$$W'' = (\theta+1)y + \mu(1+\theta-R)y - xR', \qquad \text{if } \theta > \theta'$$

The expected utility is simply

$$V = \int_{\theta'}^{\infty} U(W'')f(\theta)d\theta$$

with budget constraint $W^0=y-x$ and Lagrangian $L=V+\lambda(W_0+x-y)$.

Continuing the assumption of stochastic constant return to scale, Smith arrived at the conclusion by again taking for granted $dR/d\mu=0$:

$$V_\mu^0 = dV^0/d\mu = y^0 J(R'-R)/(1+\mu) \gtreqless 0 \quad \text{as } R'-R \gtreqless 0$$

$$\text{where } J = \int_{\theta'}^{\infty} U'(W'')f(\theta)d\theta > 0$$

Hence, his Theorem 3 (p. 72):

> Under stochastic constant returns to scale the investor whose optimal portfolio requires him to borrow on personal account to buy shares, and who is not rationed by his creditor, will prefer the corporation to increase, leave unchanged, or decrease its debt-equity ratio according as his personal borrowing rate is greater than, equal to, or less than the lending rate on corporate bonds.

But $dR/d\mu>0$. Instead, we obtain

$$\begin{aligned} (16) \quad dV^0/d\mu &= y^0 I_2 - y^0 J\mu(dR/d\mu) \\ &= y^0 J[(R'-R)/(1+\mu) - \mu(dR/d\mu)] \end{aligned}$$

upon substituting

$$\begin{aligned} I_2 &= \int_{\theta'}^{\infty} U'(W'')[(X/M) - R]f(\theta)d\theta \\ &= (\lambda - RJ)/(1+\mu) \end{aligned}$$

$$\text{and} \quad \lambda = R'J$$

From (16), we conclude $dV^0/d\mu \gtreqless 0$ according as $R'-R \gtreqless \mu(1+\mu)(dR/d\mu)$, which reduces to Theorem 3 if $dR/d\mu=0$. Since $dR/d\mu>0$, a margin investor's preference for μ depends a great deal on the current value of μ (and, of course X^*), which in turn affects R and $dR/d\mu$. For a given personal borrowing cost R', for example, an inequality $R'-R>\mu(1+\mu)(dR/d\mu)$ may be easily reversed as μ and R respond to an increase in X^*, provided that $dR/d\mu$ does not decrease sufficiently to offset the leverage effects.

REFERENCES

U. **Ben-Zion** and M. **Balch**, "Corporate Financial Theory Under Uncertainty: A Comment," *Quart. J. Econ.*, May 1973, *87*, 290–95.

J. **Hadar** and W. R. **Russell**, "Rules for Ordering Uncertain Prospects," *Amer. Econ. Rev.*, Mar. 1969, *59*, 25–34.

J. **Lintner**, "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *Rev. Econ. Statist.*, Feb. 1965, *47*, 13–37.

F. **Modigliani** and M. **Miller**. "The Cost of Capital, Corporate Finance and the Theory of Investment," *Amer. Econ. Rev.*, June 1958, *48*, 261–97.

J. **Mossin**, "Security Pricing and Investment Criteria in Competitive Markets," *Amer. Econ. Rev.*, Dec. 1969, *59*, 749–56.

V. L. **Smith**, "Corporate Financial Theory Under Uncertainty," *Quart. J. Econ.*, Aug. 1970, *84*, 451–71.

———, "Default Risk, Scale, and the Homemade Leverage Theorem," *Amer. Econ. Rev.*, Mar. 1972, *62*, 66–76.

NOTES

EIGHTY-EIGHTH ANNUAL MEETING OF THE AMERICAN ECONOMIC ASSOCIATION

Dallas, Texas, December 27–30, 1975

Preliminary Announcement of the Program

Saturday, December 27, 1975

10:00 A.M. EXECUTIVE COMMITTEE MEETING

Sunday, December 28, 1975

8:30 A.M. SAVINGS BEHAVIOR: NEW INFLUENCES AND CONSEQUENCES
Chairperson: ROBERT FERBER, University of Illinois
Papers: MARTIN FELDSTEIN, Harvard University
The Impact of Social Security
ALAN BLINDER, Princeton University
Wealth Distribution and Intertemporal Wealth Transfers
Discussants: FRANCO MODIGLIANI, Massachusetts Institute of Technology
ROBERT LAMPMAN, University of Wisconsin, Madison

8:30 A.M. UNCERTAINTY AND INCOMPLETE MARKETS*
Chairperson: DAVID CASS, University of Pennsylvania
Papers: MICHAEL ROTHSCHILD, Princeton University
Non-Optimizing Models of Administrative Behavior
HAYNE LELAND, University of California, Berkeley
Corporate Decision Making in Incomplete Markets
ROBERT FORSYTE, Carnegie-Mellon University
Efficiency in Incomplete Markets
JAMES JORDON, University of Pennsylvania
The Predictability of Prices in a Monetary Economy
Discussants: (To be announced)

8:30 A.M. ECONOMIC JUSTICE THROUGH REDISTRIBUTION: REALITY OR MYTH?* (Joint Session with the Association for the Study of the Grants Economy)
Chairperson: JULIUS REZLER, Loyola University, Chicago
Papers: TAPAN MUNROE, University of the Pacific and Massachusetts Institute of Technology, AND THEODORE TSUKAHARA, Pomona College
The Grants Economy and Contemporary Theories of Distributive Justice
SAMUEL M. LOESCHER, Indiana University
The Economics of Love and the Price System
C. LOWELL HARRISS, Columbia University
Corporate Giving: the Directions of Redistribution
RICHARD F. DYE, University of Michigan
Estimating Demand Equations for Charitable Giving
Discussants: JULIUS MARGOLIS, University of Pennsylvania
FREDERICK L. PRYOR, Swarthmore College
HOWARD P. TUCKMAN, Florida State University

8:30 A.M. FINANCE AND INDUSTRIAL ORGANIZATION* (Joint Session with the American Finance Association)
Chairperson: (To be announced)
Papers: CLEMENT G. KROUSE, University of California, Los Angeles
Capital Market and Allocative Efficiency with Technological Uncertainty
GERSHON MANDELKER, Carnegie-Mellon University
An Empirical Analysis of the Gains from Mergers
JAMES ELLERT, Queen's University
Antitrust Enforcement and the Behavior of Stock Prices

---

* Not to appear in *Papers and Proceedings*

Discussants: JOHN LONG, University of Rochester
J. FRED WESTON, University of California, Los Angeles
(Others to be announced)

10:30 A.M. CURRENT ISSUES IN METHODOLOGY*
Chairperson: RICHARD M. CYERT, Carnegie-Mellon University
Papers: LEWIS E. HILL AND GENE USELTON, Texas Tech University
Ludwig von Mises' Epistemology: A Critical Evaluation
EUGENE ROTWEIN, Queens College of the City University of New York
A Survey of the Methodological Discussion of the Last Few Decades
STEPHEN T. WORLAND, Notre Dame University
Methodological Problems of Welfare Economics: A Reflective View
Discussants: MARTIN BRONFENBRENNER, Duke University
EMILE GRUNBERG, University of Akron

10:30 A.M. THE PRICING OF OPTIONS* (Joint Session with the American Finance Association)
Chairperson: FISCHER BLACK, Massachusetts Institute of Technology
Papers: ROBERT MERTON, Massachusetts Institute of Technology
Option Pricing for Noncontinuous Stock Price Movements
JOHN COX, Stanford University
The Valuation of Financial Claims
MYRON S. SCHOLES, University of Chicago
Taxes and the Pricing of Options
HENRY LATANÉ AND RICHARD RANDLEMAN, University of North Carolina
Implied Versus Actual Standard Deviations of Daily Stock Price Ratios
Discussant: STEPHEN ROSS, University of Pennsylvania

10:30 A.M. IS THERE A CAPITAL SHORTAGE?* (Joint Session with the American Finance Association)
Chairperson: ERNEST T. BAUGHMAN, Federal Reserve Bank of Dallas
Papers: LEONALL ANDERSEN, Federal Reserve Bank of St. Louis
Capital Shortage? Theory and Empirical Evidence
ANDREW F. BRIMMER, Harvard University
Capital Markets and Competition for Long-Term Funds
ROBERT LINDSAY, New York University
Capital Shortages—Sectoral and Global
Discussants: BENJAMIN FRIEDMAN, Harvard University
SALLY RONK, U.S. Treasury Department
CHARLES UPTON, University of Chicago

10:30 A.M. NEW DEVELOPMENTS IN THE THEORY OF MONOPOLISTIC COMPETITION
Chairperson: (To be announced)
Papers: GERALD BUTTERS, Princeton University
Survey of Advertising and Market Structures
MICHAEL PORTER, Harvard University
Market Structure and Performance
MICHAEL SPENCE, Stanford University
Monopolistic Competition

10:30 A.M. THE INTERNATIONAL MONETARY MECHANISM: RECENT PROBLEMS AND EVIDENCE
Chairperson: HOLLIS CHENERY, International Bank for Reconstruction and Development
Papers: JÜRG NIEHANS, Johns Hopkins University
Current Problems with the International Monetary Mechanism
JAGDISH BHAGWATI, Massachusetts Institute of Technology
The Less Developed Countries and International Monetary Mechanism
STEPHEN MAGEE, University of Chicago
Empirical Evidence on the International Economic System
Discussants: HOSSEIN ASKARI, University of Texas at Austin
POLLY REYNOLDS ALLEN, Princeton University

---

* Not to appear in *Papers and Proceedings*

10:30 A.M. STATE AND LOCAL PUBLIC FINANCE (Joint Session with Public Choice Society and Econometric Society)
Chairperson: WILLIAM S. VICKREY, Columbia University
Papers: RICHARD E. WAGNER, Virginia Polytechnic and State University
Institutional Constraints and Local Community Formation
EITAN BERGLAS, University of Rochester and Tel-Aviv University
On the Theory of Clubs
WILLIAM C. WHEATON, Massachusetts Institute of Technology
Alternate Institution for Collective Choice Making
Discussants: RONALD E. GRIESON, Columbia University
PETER B. MIESZKOWSKI, Houston University

10:30 A.M. NATIONAL FOOD POLICY FOR AN INTERDEPENDENT AND UNCERTAIN WORLD* (Joint Session with American Agricultural Economics Association)
Chairperson: LUTHER TWEETEN, Oklahoma State University
Papers: HENDRICK S. HOUTHAKKER, Harvard University
Do Consumers or the Nation Need a National Food Policy?
GARY L. SEEVERS, Council of Economic Advisors
Impact of Policy Alternatives on the Food Sector and Its Performance
Discussants: DARYLL E. RAY, Oklahoma State University
HOWARD W. HJORT, Schnittker Associates, Washington, D.C.

10:30 A.M. APPLICATION OF OPTIMAL CONTROL TO PROBLEMS OF ECONOMIC STABILIZATION (Joint Session with the Econometric Society)
Chairperson: JAMES S. DUESENBERRY, Harvard University
Papers: GREGORY C. CHOW, Princeton University
Control Methods for Macroeconomic Policy Analysis
ALBERT ANDO, University of Pennsylvania, AND CARL PALASH, Federal Reserve Bank of New York
An Econometric Analysis of Stabilization Policies: A Use of an Optimal Control Program
JOHN H. KALCHBRENNER AND PETER A. TINSLEY, Board of Governors of the Federal Reserve System
On the Use of Optimal Control Concepts in the Design of Aggregate Monetary Policies
Discussants: HENRY C. WALLICH, Board of Governors of the Federal Reserve System
GARY FROMM, National Bureau of Economic Research

12:30 P.M. JOINT LUNCHEON (With the American Finance Association)
Chairperson: FRANCO MODIGLIANI, Massachusetts Institute of Technology
Speaker: GUIDO CARLI, Governor, Bank of Italy

2:00 P.M. INDEXATION—MONETARY CORRECTION? (Joint Session with Econometric Society and the American Finance Association)
Chairperson: STANLEY FISCHER, Massachusetts Institute of Technology
Papers: DAVID LEVHARI AND NISSAN LIVIATAN, Hebrew University
Risk and Theory of Indexed Bonds
WILLIAM POOLE, Brown University
Indexing and the Capital Markets
HENRY AARON, The Brookings Institution
Indexation of the Tax System
Discussants: FISCHER BLACK, Massachusetts Institute of Technology
KAREN JOHNSON, Stanford University
JOHN BOSSONS, University of Toronto

2:00 P.M. INCENTIVES UNDER SOCIALISM: MODELS AND REALITY*
Chairperson: ZBIGNIEW M. FALLENBUCHL, University of Windsor
Papers: LEONID HURWICZ, University OF Minnesota
Incentives and Efficiency in Economic Decisions
ARON KATSENELINBOIGEN AND HERBERT S. LEVINE, University of Pennsylvania
Incentive Mechanisms in the Soviet Economy
C. ROBERT ROLL, The RAND Corporation
Incentives and Motivation in China
Discussants: EGON NEUBERGER, State University of New York at Stony Brook
FREDERIC M. SCHERER, Federal Trade Commission
ROBERT F. DERNBERGER, University of Michigan

---

* Not to appear in *Papers and Proceedings*

2:00 P.M. GRANTS IN MONETARY POLICY AND THE STRUCTURE OF THE FINANCIAL SYSTEM* (Joint Session with the Association for the Study of the Grants Economy)
Chairperson: ANDREW F. BRIMMER, Harvard University
Papers: GEORGE M. VON FURSTENBERG, Council of Economic Advisors
An Analytical Framework for Identifying Subsidy Components in Government Sponsored Credit
RICHARD COHN, Massachusetts Institute of Technology
The Implications for Mortgage Lenders of Removing Deposit Rate Ceilings
PATRIC H. HENDERSCHOTT, Purdue University
Grant Elements in the Operation of Federally Sponsored Credit Agencies in the Mortgage Market
THOMAS F. WILSON, Butler University
Identification and Measurement of Grant Elements in Monetary Policy
Discussants: JOSEPH SCHERER, Federal Reserve Bank of New York
THOMAS HAVRILESKY, Duke University

2:00 P.M. EXPERIMENTAL ECONOMICS
Chairperson: MARTIN SHUBIK, Yale University
Papers: JAMES FRIEDMAN, University of Rochester, AND AUSTIN HOGGATT, University of California, Berkeley
Pricing Signalling in Experimental Oligopoly
ROGER NOLL, California Institute of Technology
On the Pricing of Public Goods: Public Television Programming
VERNON L. SMITH, University of Arizona
Experimental Economics: Some Theories and Results
Discussants: CHARLES PLOTT, California Institute of Technology
MICHAEL ROTHSCHILD, Princeton University

2:00 P.M. MARXIST PERSPECTIVES ON THE CAPITALIST MACROECONOMY
Chairperson: LEONARD A. RAPPING, University of Massachusetts
Papers: JENS CHRISTIANSEN, Stanford University
Marx and the Falling Rate of Profit
JAMES CROTTY, University of Massachusetts, AND RAYFORD BODDY, American University
Capitalism, Instability, and Economic Policy
ARTHUR MACEWAN, Harvard University
World Capitalism and U.S. Economic Instability
Discussants: GEORGE FEIWEL, University of Tennessee
DONALD J. HARRIS, University of West Indies

2:00 P.M. NEW DEVELOPMENTS IN LABOR MARKET DYNAMICS (Joint Session with Econometric Society and Industrial Relations Research Association)
Chairperson: EDWARD KALACHEK, Washington University
Papers: CHARLES C. HOLT, RALPH E. SMITH, RICHARD S. TOIKKA, AND WILLIAM J. SCANLON, The Urban Institute
A Simulation Model of the Demographic Composition of Employment, Unemployment and Labor Force Participation
ROBERT J. FLANAGAN, Stanford University
The Economics of Racial Unemployment Differentials
ROBERT HALL AND RICHARD KASTEN, Massachusetts Institute of Technology
Labor Mobility and Occupational Success of Young Men
Discussants: FRANK BRECHLING, Northwestern University
CLIFFORD REID, Grinnell College
EMILY ANDREWS, U.S. Department of Labor–ASPER

2:00 P.M. ECONOMIC FLUCTUATIONS AND STABILIZATION POLICY, 1965–75: SOME ECONOMETRIC EVIDENCE*
Chairperson: LAWRENCE R. KLEIN, University of Pennsylvania
Papers: ALBERT A. HIRSCH, U.S. Department of Commerce
Econometric Review of Alternative Fiscal and Monetary Policies 1965–75—I
HAROLD SHAPIRO, University of Michigan
Econometric Review of Alternative Fiscal and Monetary Policies 1965–75—II
KEITH JOHNSON, Columbia University
Econometric Review of Alternative Fiscal and Monetary Policies, International Aspects—I

* Not to appear in *Papers and Proceedings*

Discussants: HENRY C. WALLICH, Board of Governors of the Federal Reserve System
WILLIAM FELLNER, American Enterprise Institute
R. JEFFREY GREEN, Council of Economic Advisers
GEORGE D. N. WORSWICK, National Institute of Economic and Social Research

2:00 P.M. POLITICAL ECONOMY: SOME USES OF THE EXIT-VOICE APPROACH
Chairperson: JAMES TOBIN, Yale University
Papers: RICHARD B. FREEMAN, Harvard University
Individual Mobility and Collective Action in the Labor Market
OLIVER E. WILLIAMSON, University of Pennsylvania
The Economics of Internal Organizations: Exit and Voice in Relation to Markets and Hierarchies
DENNIS R. YOUNG, State University of New York at Stony Brook
Consolidation or Diversity: Choices in the Structure of Urban Governance
Discussants: ALBERT O. HIRSCHMAN, Institute for Advanced Study
RICHARD R. NELSON, Yale University

8:00 P.M. RICHARD T. ELY LECTURE
Chairperson: JAMES TOBIN, Yale University
Speaker: ASSAR LINDBECK, Institute for International Economic Studies
Stabilization Policy in Open Economies with Endogenous Politicians

Monday, December 29, 1975

8:30 A.M. ENVIRONMENTAL ECONOMICS* (Joint Session with the American Association of Environmental Economists)
Chairperson: RALPH C. D'ARGE, University of Wyoming, AND ROBERT DORFMAN, Harvard University
Papers: WILLIAM J. BAUMOL, Princeton University
Environmental Economics in Perspective
A. NICHOLS AND RICHARD ZECKHAUSER, Harvard University
Should the Federal Government Stockpile Resources?
CLIFFORD RUSSELL AND WALTER SPOFFORD, Resources for the Future, Inc.
A Regional Environmental Management Model: Some Results
M. L. CROPPER, University of California, Riverside
Regulating Activities with Catastrophic Environmental Effects
Discussants: (To be announced)

8:30 A.M. ECONOMIC ISSUES IN IMPOSING MULTIPLE OBJECTIVES IN THE TRANSPORT-UTILITY SECTOR* (Joint Session with the Transportation and Public Utilities Group)
Chairperson: WILLIAM H. DODGE, University of Wisconsin
Papers: ROBERT W. HARBESON, University of Illinois
Social Welfare and Economic Efficiency in Transport Policy
TILLO E. KUHN, York University
Economies and Multiple Objectives in Urban Transportation
HARRY M. TREBING, Michigan State University
Broadening the Objectives of Public Utility Regulation
Discussant: JOEL DIRLAM, University of Rhode Island

8:30 A.M. SELECTION AND DIRECTION OF JOURNAL MATERIAL WITH REGARD TO CAREER DEVELOPMENT*
Chairperson: MARK PERLMAN, University of Pittsburgh
Papers: WILLIAM BRAINARD, Yale University
Some Results of the American Economic Association Readership Survey
HUGH NORTON, University of South Carolina
Equity and Effectiveness of the Editing Process as Perceived by the Producers of Manuscripts
ALICE VANDERMEULEN, University of California, Los Angeles
The Publication Problem as Viewed by an Erstwhile Editor
ROBERT EAGLY, Federal Reserve Bank of New York
Measuring the Scholarly Effectiveness of Journal Articles
Discussants: CHARLES WILSON, University of California, Los Angeles
MICHAEL LOVELL, Wesleyan University
DAVID A. KATZ, University of Missouri, Rolla

---

* Not to appear in *Papers and Proceedings*

8:30 A.M. IMPROVING THE DATA BASE FOR THE U.S. NATIONAL INCOME AND PRODUCT ACCOUNTS* (Joint Session with the Econometric Society)
Chairperson: David T. Hulett, Statistical Policy Division, Office of Management and Budget
Panelists: ROSANNE E. COLE, International Business Machines
ROBERT J. GORDON, Northwestern University
DALE W. JORGENSEN, Harvard University
DANIEL CREAMER, Consultant to the Conference Board and Chairman of GNP Data Improvement Project
EDWARD DENISON, The Brookings Institution

8:30 A.M. INFLATION AND STRUCTURAL ADJUSTMENTS IN THE RESIDENTIAL MORTGAGE MARKET* (Joint Session with the American Finance Association and American Real Estate and Urban Economics Association)
Chairperson: JACK GUTTENTAG, University of Pennsylvania
Papers: RICHARD COHN AND DONALD LESSARD, Massachusetts Institute of Technology
Recent Research on Indexation and the Housing Market
DONALD TUCKER, Board of Governors, Federal Reserve System
Indexation and the Housing Market
STANLEY S. STANSELL AND JAMES MILLER, Farm Credit Association
Study of Variable Rate Farm Credit Residential Mortgages
Discussants: ROBERT EDELSTEIN, University of Pennsylvania
RICHARD V. EASTIN, University of Southern California
ALAN DONZIGER, Villanova University

10:30 A.M. RECENT DEVELOPMENTS IN THE APPLICATION TO ECONOMICS OF DIFFERENTIAL TOPOLOGY AND GLOBAL ANALYSIS (Joint Session with Econometric Society)
Chairperson: TRUMAN BEWLEY, Harvard University
Papers: GERARD DEBREU, University of California, Berkeley
Regular Differentiable Economies
STEPHEN SMALE, University of California, Berkeley
Global Analysis and Dynamics in Equilibrium Theory
Discussants: DONALD BROWN, Yale University
HAL VARIAN, Massachusetts Institute of Technology

10:30 A.M. INFORMATION AND MARKET STRUCTURE (Joint Session with the Econometric Society)
Chairperson: VERNON L. SMITH, University of Arizona
Papers: STEVEN SALOP, Federal Reserve Board
Information and Industrial Organization
JOSEPH E. STIGLITZ, Stanford University
Information, and Capital Market: A Survey
JOHN RILEY, University of California, Los Angeles
Information, Screening, and Human Capital
Discussants: PAUL COOTNER, Stanford University
ELIZABETH E. BAILEY, Bell Laboratories and New York University
DALE MORTENSEN, Northwestern University

10:30 A.M. INTERGENERATIONAL TRANSFERS OF INEQUALITY (Joint Session with the Econometric Society)
Chairperson: ORLEY ASHENFELTER, Princeton University
Papers: MARY CORCORAN, Michigan State University, AND CHRISTOPHER JENCKS, Harvard University
The Effects of Family Background on Income: A Review and Some New Evidence
JERE BEHRMAN AND PAUL TAUBMAN, University of Pennsylvania
Intergenerational Transfers of Earnings and Occupational Status: The Contributions of Genetics and Environment
GLEN LAURI, Massachusetts Institute of Technology
Equal Employment Opportunity and Racial Income Differences
Discussants: RICHARD B. FREEMAN, Harvard University
MICHAEL L. WACHTER, University of Pennsylvania
FINIS WELCH, University of California, Los Angeles

---

* Not to appear in *Papers and Proceedings*

10:30 A.M. NEW DEVELOPMENTS IN PUBLIC FINANCE (Joint Session with the Econometric Society)
Chairperson: MARTIN FELDSTEIN, Harvard University
Papers: DAVID BRADFORD, Princeton University
Optimal Taxation
MICHAEL BOSKIN, Stanford University
Recent Econometric Research on Taxation
Discussants: EYTAN SHESHINSKI, Hebrew University
GEORGE BREAK, University of California, Berkeley

10:30 A.M. RISK BENEFIT ANALYSIS OF ALTERNATIVE ENERGY SOURCES*
Chairperson: ALAN S. MANNE, Harvard University
Papers: LESTER LAVE, Carnegie-Mellon University
Coal or Nuclear: The External Costs of Electricity Generation
THOMAS COCHRAN, Natural Resources Defense Council, Inc.
Risk Benefit Analysis in Nuclear Decisions
SAUL LEVINE, Nuclear Regulatory Commission
Overview of Nuclear Power Plant Risk Assessment
Discussants: (To be announced)

10:30 A.M. INCOME DISTRIBUTION IN DEVELOPING COUNTRIES (Joint Session with the Association for Comparative Economic Studies)
Chairperson: PAUL ROSENSTEIN-RODAN, Boston University
Papers: IRMA ADELMAN, University of Maryland, AND SHERMAN ROBINSON, Princeton University
Income Distribution and Structural Change in Developing Countries
MONTEK AHLUWAHLIA, International Bank for Reconstruction and Development
Income Distribution and Economic Development: Some Stylized Facts
MICHAEL BEST, University of Massachusetts
Uneven Development and Dependent Market Economies
Discussants: WILLIAM R. CLINE, The Brookings Institution
LANCE TAYLOR, Massachusetts Institute of Technology
ALEX O. WILLIAMS, University of Pittsburgh

10:30 A.M. CURRENT ISSUES IN THE FINANCING OF DEVELOPMENT*
Chairperson: ISAIAH FRANK, Johns Hopkins University
Papers: RAYMOND F. MIKESELL, University of Oregon
External Finance and the Current Resource Transfer Theory
JORGE RUIZ-LARA AND NICHOLAS BRUCK, Inter-American Development Bank
Development Finance Requirements at Different Stages of Development
WOUTER TIMS AND JO W. SAXE, International Bank for Reconstruction and Development
The Role of Petro-Dollars and Euro-Currencies in Development Finance
Discussants: HOSSEIN ASKARI, University of Texas at Austin
LESTER GORDON, Harvard Institute for International Development
ALBERT O. HIRSCHMAN, Institute for Advanced Study, Princeton University

2:00 P.M. TREND AND ANALYSIS IN BLACK ECONOMIC DEVELOPMENT (Joint Session with the National Economic Association)
Chairperson: ALFRED L. EDWARDS, University of Michigan
Papers: WILLIAM D. BRADFORD, Stanford University, AND ALFRED OSBORNE, University of California, Los Angeles
A Theoretical Investigation of the Entrepreneurship Decision in the Context of Black Economic Development
WALTER W. MCMAHON, University of Illinois
Influences on Investment by Blacks in Higher Education
KARL D. GREGORY, Oakland University, AND FLOYD HOGAN, Black Economic Research Center, New York
Recent Trends in the Development of Minority-Owned and Operated Businesses
KENNETH KNIGHT, University of Texas at Austin
Capital Problems in Minority Business Development
Discussants: ERIC SCHENKER, University of Wisconsin at Milwaukee
VINCENT R. MCDONALD, Howard University
THERESA SHAPIRO, University of New Orleans

* Not to appear in *Papers and Proceedings*

2:00 P.M. THE FLEXIBILITY OF WAGES AND PRICES
Chairperson: ROBERT J. GORDON, Northwestern University
Papers: GEORGE AKERLOF, University of California, Berkeley
Inflation: Tales Told by Static Models
WILLIAM D. NORDHAUS, Yale University
The Sources of Worldwide Inflation
MICHAEL L. WACHTER, University of Pennsylvania
The Flexibility of Wages in Postwar Recessions
Discussants: ASSAR LINDBECK, Institute for International Economic Study
DAVID MEISELMAN, Virginia Polytechnic Institute at Reston
DAVID H. SWINTON, State University of New York at Stony Brook

2:00 P.M. INTERNATIONAL ECONOMIC INTERDEPENDENCE: IMPLICATIONS FOR POLICY
Chairperson: ANNE O. KRUEGER, University of Minnesota
Papers: ROBERT BALDWIN, University of Wisconsin
The Effect of Alternative Tax Cutting Rules on the Industrial Composition of U.S. Trade and Employment
THOMAS HORST, The Brookings Institution and Harvard University
Multinational Firms and National Taxation
C. FRED BERGSTEN AND WILLIAM R. CLINE, The Brookings Institution
Implications of Growing Interdependence for Research Agenda
Discussants: GOTTFRIED HABERLER, American Enterprise Institute
HARRY G. JOHNSON, University of Chicago

2:00 P.M. RECENT AND PROSPECTIVE CHANGES IN THE STRUCTURE OF METROPOLIS* (Joint Session with the Real Estate and Urban Economics Association)
Chairperson: CHARLES L. LEVEN, Washington University
Papers: WILLIAM ALONSO, University of California, Berkeley
Demographic Shifts and Urban Structures
ANN SCHNARE, The Urban Institute
Energy, Transportation, and Urban Structure
EDWIN S. MILLS, Princeton University
Land Use Controls and Energy Costs in Urban Areas
JEROME ROTHENBERG, Massachusetts Institute of Technology
Impact of Energy Costs on Intra-Urban Location
Discussant: EUGENE SMOLENSKY, University of Wisconsin

2:00 P.M. ECONOMIC STATUS OF WOMEN IN INTERNATIONAL WOMEN'S YEAR
Chairperson: BARBARA B. REAGAN, Southern Methodist University
Papers: MARY KEYSERLING, Washington, D.C.
National Aspects
(To be announced)
International Aspects
Discussants: NANCY BARRETT, Congressional Budget Office
KENNETH E. BOULDING, University of Colorado
JUANITA KREPS, Duke University

2:00 P.M. EXCHANGE VERSUS GRANT TRANSACTIONS IN ENVIRONMENTAL MODELS (Joint Session with the Association for the Study of the Grants Economy)
Chairperson: KENNETH E. BOULDING, University of Colorado
Papers: MARTIN PFAFF, University of Augsburg, FRG
The Distribution Effects of the Polluter Pays Principle
LLOYD D. ORR, Indiana University
Incentive for Innovation as the Criterion for Effluent Charges
HAROLD BRUMM, California State University, Los Angeles
The Role of Government in the Allocation of Resources to Technological Innovation
RON D. WHITE, Wichita State University
The Anatomy of Nonmarket Failure: A Structure for Analyzing Policy Decisions
Discussants: A. ALLAN SCHMID, Michigan State University
JANOS HORVATH, Holcomb Research Institute, Butler University

---

* Not to appear in *Papers and Proceedings*

2:00 P.M. HISTORY OF POLITICAL ECONOMY AND BIO-ECONOMICS* (Joint Session with the History of Economics Society)
Chairperson: CARL G. UHR, University of California, Riverside and Vice President of the History of Economics Society
Papers: NICHOLAS GEORGESCU-ROEGEN, Vanderbilt University
Economics or Bio-Economics?
WILLIAM D. GRAMPP, University of Illinois at Chicago Circle
Economists and Politicians: Some Cautionary History
Discussants: ROBERT M. LARSEN, University of Utah
JOAN MCCREA, University of Texas at Arlington

2:00 P.M. CURRENT ISSUES IN DEFENSE ECONOMY*
Chairperson: JOHN KOEHLER, Congressional Budget Office
Papers: JOHN DEPRES, The RAND Corporation
Political Economy of Modern Military Competition
MICHAEL REICH, University of California, Berkeley
Obstacles to Controlling Military Spending
BYRON DOENGES, Arms Control and Disarmament Agency
International Trade in Arms
Discussants: JOHN STEINBRUNER, John F. Kennedy School of Government
ROBERT L. SLIGHTON, U.S. Treasury

2:00 P.M. HOW CAN GOVERNMENT NEEDS FOR POLICY RESEARCH BE BETTER MET?*
Chairperson: CHARLES C. HOLT, The Urban Institute
Panelists: ALICE M. RIVLIN, Congressional Budget Office
CHARLES L. SCHULTZE, The Brookings Institution
MICHAEL H. MOSKOW, Department of Housing and Urban Development
ERWIN HARGROVE, The Urban Institute and Brown University
WILLIAM NISKANEN, University of California, Berkeley

2:00 P.M. THE ECONOMIC ORGANIZATION OF SOCIALIST AGRICULTURE (Joint Session with the Association for Comparative Economic Studies and the American Agricultural Economics Association)
Chairperson: ARCADIUS KAHAN, University of Chicago
Papers: D. GALE JOHNSON, University of Chicago
An Economic Model for Socialist Agriculture
ANTHONY TANG, Vanderbilt University
Organization and Performance of Chinese Agriculture
ZBIGNIEW FALLENBUCHL, University of Windsor
Incentives and the Organization of East-European Agriculture
ROY LAIRD, University of Kansas
Central Controls, Farm Size and the Zveno
ROBERT STUART, Douglass College
The Socialist Organizational Model and the Developing Economies

4:00 P.M. BUSINESS MEETING

8:00 P.M. PRESIDENTIAL ADDRESS
Chairperson: GEORGE LELAND BACH, Stanford University
Speaker: ROBERT AARON GORDON, University of California, Berkeley

Tuesday, December 30, 1975

8:30 A.M. IN ANTICIPATION OF NATIONAL HEALTH INSURANCE: ANALYSES OF HEALTH COST CONTAINMENT PROGRAMS* (Joint Session with the Health Economics Research Organization)
Chairperson: DONALD E. YETT, University of Southern California
Papers: CLIFTON R. GAUS AND BARBARA S. COOPER, Social Security Administration
Evaluation of Medicaid Experience in Health Maintenance Organizations
WILLIAM D. FINKLE, Southern California Kaiser Foundation and Health Plan
Quality of Care Versus Cost Containment: Implication for Professional Standards Review Organizations (PSROs)
RALPH E. BERRY, JR., Harvard University
Evaluation of Prospective Hospital Reimbursement in New York
Discussants: PAUL J. FELDSTEIN, University of Michigan
ROYAL A. CRYSTAL, Bureau of Quality Assurance, Department of Health, Education, and Welfare
JUDITH R. LAVE, Carnegie-Mellon University

* Not to appear in *Papers and Proceedings*

8:30 A.M. PLANNING FOR THE 1980 CENSUS*
Chairperson: RICHARD RUGGLES, Yale University
Papers: DAVID L. KAPLAN, Bureau of the Census
Overall Issue and Problems
MEYER ZITTER, Chief of the Population Division
Subject Content of the Population Census
ARTHUR F. YOUNG, Chief of the Housing Division
Subject Content of the Housing Census

8:30 A.M. ALLOCATIONAL EFFICIENCY OF SECURITY MARKETS UNDER UNCERTAINTY* (Joint Session with the American Finance Association)
Chairperson: ROBERT LITZENBERGER, Stanford University
Papers: SANFORD GROSSMAN, Stanford University
Incentives for Information Gathering in Competitive Stock Markets where Investors have Diverse Knowledge
NEILS NIELSEN, Copenhagen School of Economics and Business Administration
The Investment Decision of the Firm Under Uncertainty
STEPHEN ROSS, University of Pennsylvania
Behavior of Agents, Efficiency and Uncertainty
MARK RUBINSTEIN, University of California, Berkeley
The Strong Case for the Generalized Logarithmic Utility Model as the Premier Model of Financial Markets
Discussants: ALAN KRAUS, University of Washington
HAYNE LELAND, University of California, Berkeley
JOSEPH STIGLITZ, Stanford University
FRED GRAUER, Massachusetts Institute of Technology

8:30 A.M. ECONOMIC EDUCATION
Chairperson: PHILLIP SAUNDERS, University of Indiana
Papers: ELISABETH ALLISON, Harvard University
Self-Paced Student Learning at Harvard
DARRELL LEWIS AND WILLIAM BECKER, University of Minnesota
The Teacher Training Program for New Ph.D.s
STEPHEN LONG, Franklin and Marshall College, AND RICHARD HANSEN, University of Nebraska
Participants' Views of the Teacher Training Program
Discussants: JOHN SIEGFRIED, Vanderbilt University
CAMPBELL McCONNELL, University of Nebraska
DAWN ELVIS, Fisk University

8:30 A.M. DYNAMICS OF THE POVERTY AND WELFARE POPULATION*
Chairperson: HEATHER ROSS, Office of Economic Analysis, U.S. Department of the Interior
Papers: DAVID LYON AND OTHERS, staff members, The New York City Rand Institute
A Dynamic Analysis of the New York City Welfare Rolls
MICHAEL WISEMAN, University of California, Berkeley
The Dynamics of the Alameda County, California Welfare Rolls
BARRY L. FRIEDMAN AND LEONARD J. HAUSMAN, Brandeis University
Work and Welfare Patterns in Low Income Families
FRANK LEVY, University of California, Berkeley
How Big is the American Underclass?
Discussants: HENRY AARON, The Brookings Institution
MICHAEL BARTH, University of Wisconsin and Department of Health, Education, and Welfare

10:30 A.M. THE LAST GREAT DEPRESSION AND THE PRESENT ONE
Chairperson: ANNA J. SCHWARTZ, National Bureau of Economic Research
Papers: PETER TEMIN, Massachusetts Institute of Technology
Lessons for the Present from the Great Depression
ARTHUR GANDOLFI AND JAMES R. LOTHIAN, First National City Bank in New York
Depression, Uncertainty, and the Demand for Money
HEYWOOD FLEISIG, Federal Reserve System
The International Recycling Problem: the Interwar Years and Today

---

* Not to appear in *Papers and Proceedings*

Discussants: Robert A. Gordon, University of California, Berkeley
Sidney Davis, Atlanta University
Michael R. Darby, University of California, Los Angeles

10:30 A.M. Economics of Information and Innovation*
Chairperson: Joel Goldhar, National Science Foundation
Papers: William Baumol, Princeton University and New York University
Public Good Attributes of Information and its Optional Pricing
Fritz Machlup, New York University
The Economic Strangulation of the Science Information Enterprise
Edwin Mansfield, University of Pennsylvania
Returns from New Industrial Technology
Discussants: Sanford Berg, University of Florida
Nestor Terleckyj, National Planning Association
Robert Willig, Bell Telephone Laboratories

10:30 A.M. International Capital Markets and Foreign Exchange: New Directions of Research (Joint Session with the American Finance Association)
Chairperson: (To be announced)
Papers: Kalman J. Cohen and others, Duke University and New York University
The Determinants of Common Stock Returns Volatility: An International Comparison
Michael Adler and Bernard Dumas, Columbia University
Variations in the Demand for Stocks and Forward Currency Contracts in an International Capital Asset Market with Default Risk
Richard W. Roll and Bruno H. Solnik, Centre d'Enseignment Superieur des Affaires
Interest Rates and Exchange Rates
Discussants: Tamir Agmon, Tel Aviv University
Donald Lessard, Massachusetts Institute of Technology
Haim Levy, Jerusalem University

10:30 A.M. Expectations and Aggregate Dynamics* (Joint Session with the Econometric Society)
Chairperson: Robert E. Lucas, Jr., University of Chicago
Papers: Robert J. Barro, University of Rochester
Indexation in a Rational Expectation Model
Edmund S. Phelps and John Taylor, Columbia University
Rational Expectations: No Bar to Monetary Stabilization Policy
Thomas Sargent, University of Minnesota
Estimates of the Demand for Money in Hyper Inflation
Discussants: Stanley Fischer, Massachusetts Institute of Technology
Bennett McCallum, University of Virginia
Robert Shiller, University of Pennsylvania

10:30 A.M. New Directions in Health Policy Research* (Joint Session with the Health Economics Research Organization)
Chairperson: Victor F. Fuchs, National Bureau of Economic Research
Papers: Karen Davis, The Brookings Institution, and Ray Marshall, University of Texas
The Market for Health Care in Rural Areas
Mark Pauly, Northwestern University
The Role of Demand Creation in the Provision of Health Care Services
Uwe Reinhardt, Princeton University
Health Manpower Planning in a Market Context
Discussants: Herbert E. Klarman, New York University
Stuart Altman, U.S. Department of Health, Education, and Welfare
Barbara Kehrer, Mathematica

Note: In addition to these sessions of invited papers the American Economic Association is also arranging for several sessions of contributed papers, including some sponsored with the Econometric Society. The scheduled times and other details will be given in the final program to be distributed at the meetings.

* Not to appear in *Papers and Proceedings*

ANNOUNCEMENTS

On March 14, 1975 the Executive Committee of the American Economic Association approved the following changes in the financial operation of the *American Economic Review* and *Journal of Economic Literature.*

a) The discontinuance of free reprints/offprints to authors of material appearing in the *American Economic Review* and *Journal of Economic Literature.* Reprints/offprints will continue to be sold at prices which cover the relevant costs. This change will begin with the December 1975 issues.

b) A differential submission fee for authors who are not members of the American Economic Association. Beginning July 1, 1975, the submission fee for manuscripts submitted to the *American Economic Review* will be fifteen dollars ($15) for members and thirty dollars ($30) for nonmembers.

c) A postage surcharge for mailing of material outside the United States. This surcharge will equal the difference between mailing costs inside and outside the United States. It will be added to new subscription and membership bills January 1, 1976.

d) A page charge for articles, shorter papers, notes, comments, and replies published in the *American Economic Review* and in the *Journal of Economic Literature.* The page charge is payable by the institution or granting agency supporting the research. Payment of this charge is not a prerequisite for publication, nor are authors expected to pay these charges themselves. The charge will be twenty-five dollars ($25) per printed page, and will go into effect with manuscripts accepted after July 1, 1975.

The eighty-eighth annual meeting of the American Economic Association will be held in Dallas, Texas, December 28–30, 1975. The Employment Registry and Center will open December 27. The meeting was originally scheduled for October, but has been changed in the interest of attracting greater attendance.

Nicolaus and Associates have been appointed the official travel coordinator for the Dallas meetings of the Allied Social Science Associations. Group air fares may be arranged from your city and for groups as small as fifteen. These fares usually run about 25 percent below regular coach fare. Some examples using current fares subject to change are:

	Coach	*Group*
New York–Dallas-New York	$236.73	$159.73
Des Moines–Dallas–Des Moines	124.73	99.73
San Francisco–Dallas–San Francisco	252.73	202.73

Flights will be arranged for other cities if enough individuals show interest. Air reservation forms and general information on flights and costs will be mailed to you in October with the preregistration material. In the meantime if you have any questions on group flights, please call Nicolaus and Associates collect at 817-461-1622, or write them at 1201 N. Watson Road, Suite 179, Arlington, Texas 76011.

Five changes in the bylaws have been approved by the Executive Committee and are being submitted to the membership for possible adoption by a mail ballot. A full description of the proposed changes together with the ballot have been included in the mailing of material (including ballots) related to the election of next year's officers. Three of the changes concern the dues structure and classes of membership. On account of the poor financial condition of the Association (see this *Review*, May 1975, pages 460–63 and 467–74), increased revenues are needed. It is proposed to increase the progressivity of the dues structure, putting the burden primarily on those with higher incomes; to double the dues of family members; and to phase out the category of life members. Another proposal would give the Executive Committee the option of submitting to the entire membership for mail vote any resolution adopted at an annual meeting by a vote of less than 5 percent of the members (counting negative as well as positive votes). The fifth proposal would remove the limit of twelve to the number of elected Distinguished Fellows.

Resolutions for Consideration at the Annual Business Meeting

The Executive Committee at its meeting on March 8, 1974, voted to require that, to be considered at the annual business meeting, proposed resolutions must be submitted to the Secretary at least one month in advance in writing with signatures of the proposer and the second, both of whom must be members in good standing. The Secretary will reproduce the proposed resolutions and make copies available in advance of the meeting. The next business meeting will be on December 29, 1975. The deadline for proposed resolutions is, accordingly, November 29. They should be sent to the Secretary, American Economic Association, 1313 21st Avenue South, Nashville, Tennessee 37212.

The Asia Foundation has awarded the American Economic Association a grant to assist graduate students and visiting scholars from Asia currently at *U.S.* institutions to attend the annual meeting of the association. The 1975 meeting will be held in Dallas, Texas, December 28–30. Approximately 8 travel grants up to $150 each will be awarded. Write the American Economic Association, 1313 21st Avenue South, Nashville, Tennessee 37212 for an application form. To be considered for a travel grant, completed forms must be received by November 7, 1975.

Call for Papers for the 1976 Meetings

Members wishing to give papers or make suggestions for the program for the meetings to be held in Atlantic City, September 16–18, 1976, are invited to send their ideas to Rendigs Fels, Secretary, American Economic Association, 1313 21st Avenue South, Nashville, Tennessee 37212. Although most of the sessions sponsored by the American Economic Association will consist of invited papers, there will also be several sessions of noneconometric contributed papers. Proposals for invited sessions should be submitted as soon as possible, as this phase of program planning will be completed by mid-autumn. To be considered for the contributed sessions, abstracts of proposed (noneconometric) papers must be received by the Secretary no later than February 1, 1976. Economists wishing to give papers on econometrics or economic theory may submit abstracts to the Econometric Society, which meets with the American Economic Association and annually schedules a substantial number of contributed sessions.

Economists who are *strongly* oriented toward the humanities, who use humanistic methods in their research, and who will be participating in meetings abroad that are concerned with the humanistic aspects of their discipline are eligible to apply for travel grants of the American Council of Learned Societies. Specifically, economists may be eligible if (a) they deal with the history of economic thought or economic history and (b) if their approach is qualitative and descriptive rather than quantitative and statistical. Conferences dealing with the establishment of social policy or legislation are generally ineligible. The deadlines for applications to be received in the office of the American Economic Association are: for meetings scheduled between June and September, February 15; for meetings scheduled between October and January, June 15; for meetings scheduled between February and May, October 15. Application forms may be obtained from Rendigs Fels, Secretary, American Economic Association, 1313 21st Avenue South, Nashville, Tennessee 37212.

1976–77 FDIC Graduate Fellowships

The Federal Deposit Insurance Corporation is currently accepting applications for its 1976–77 fellowships which support research in banking, finance, and economics. Graduate students who will have completed all requirements for the Ph.D. (except the dissertation) by October 1, 1976, and are about to begin writing their doctoral dissertations in the areas of economics, banking, and finance are invited to apply for this year's fellowships. The maximum fellowship award is $6,000 (plus tuition), with an additional award of $500 if the dissertation is completed and accepted within the period covered by the fellowship. The awards cover a period of 12 to 15 months, beginning any month between July and October 1976, and ending no later than September 30, 1977. Requests for fellowship applications should be directed to: Federal Deposit Insurance Corporation, Division of Research, 550 17th Street, N.W., Washington, D. C. 20429.

The Social Science History Association has been organized to further the interests of historians and social scientists concerned with the application of social science theory and method in historical context. Members of the Association plan a wide range of activities, including a publication program, national and regional conferences, the fostering of special interest networks, data resource and curriculum development and related action. Charter, regular, and student memberships ($100, $15, and $8) are available: write Jerome M. Clubb, Executive Secretary, SSHA, Center for Political Studies, University of Michigan, P.O. Box 1248, Ann Arbor, Michigan 48106. Robert Swierenga, Department of History, Kent State University, Kent, Ohio 44240, editor of the Association Journal, is now accepting manuscripts for review for publication in the journal.

Economists spending time in Europe desiring lecturing appointments in Common Market countries are urged to write to Mr. William Zavis, Deputy Public Affairs Counselor, U.S. Mission to the European Communities, 23 Avenue des Arts, 1040 Brussels, Belgium.

The first African Conference on Informatics in Administration will be held in early 1976. Individuals interested in participating or preparing research papers to be presented should contact Klaus D. Grimm, Special Research Consultant, State of California, 745 Franklin Street, Suite 410, San Francisco, California 94102.

A bicentennial conference, "Adam Smith and the Wealth of Nations, 1776–1976," will be held at Eastern Kentucky University in the spring of 1976, tentatively March 26–28. Scholars in the fields of history, economics, philosophy, political science, and any other field touching on the wide-ranging concerns of Smith's writings are invited to submit proposals for papers. Interested persons should write to Professor William R. Morrow, Department of Economics, Eastern Kentucky University, Richmond, Kentucky 40475.

A conference, "American Re-Evolution," will be held March 19–21, 1976, at the Arizona Inn, Tucson, sponsored by the Department of Economics of the University of Arizona, Criterion Studies, and the Public Choice Society. Authors of papers which explore new ways to "govern" ourselves, or isolate causes of our current difficulties, are invited to submit copies no later than November 25, 1975, or preliminary outlines by no later than October 25, to Professor Richard D. Auster, Department of Economics, University of Arizona, Tucson, Arizona 85721. Travel and housing expenses of participants in the conference will be paid.

The second annual Virginia Conference on Urban Studies, "The Urban South—Perspective '76," will be held February 12–13, 1976 at Norfolk State College. Urban specialists from all the social sciences are invited to participate. For further information, write Professor Jesse Pendleton, Department of Geography, Norfolk State College, or Professor Carl Abbott, Department of History, Old Dominion University, Norfolk, Virginia 23508.

Under a joint program of the Social Science Research Council and the American Council of Learned Societies, fellowships are offered for nine to eighteen months of dissertation research on contemporary Western European affairs. Applications are invited from students in all social sciences and humanities, but particular attention is given to disciplines in which relatively less attention has been devoted to Western Europe, such as anthropology, economics, social psychology, and sociology. Requests for further information and application forms should be sent to Social Science Research Council, Western European Program, 605 Third Avenue, New York, New York 10016.

The Southern Regional Science Association will hold its 1975 meetings in Richmond, Virginia, on April 8–9, 1976. Scholars interested in presenting papers on any subject involving analysis of the spatial dimensions of human activity are invited to submit one-page abstracts by October 1, 1975 to the program chairman, Professor Monroe Newman, Department of Economics, Pennsylvania State University, University Park, Pennsylvania 16802.

The purpose of the History of Economics Society is to promote interest and inquiry into the history of economics and related parts of intellectual history; facilitate communication and discourse among scholars working in the field; and acquaint members of the profession with the scientific, literary, and philosophical tradition of economics. Information about membership in the Society and about a reduced-price subscription to the quarterly journal *History of Political Economy* can be obtained from James L. Cochrane, Secretary-Treasurer, History of Economics Society, University of South Carolina, Columbia, South Carolina 29208.

The Society will hold its next annual conference at Center for Continuing Education, University of Chicago on May 21–23, 1976. The primary but not sole theme of the conference will be Adam Smith's *Wealth of Nations*. Anyone wishing to present a paper should submit an abstract (two copies) together with a *separate* sheet containing the following information: Author's name, address, professional affiliation, position, and telephone number; title of paper; and whether or not the author is a member of the Society. This must be received by November 15, 1975. Persons wishing to serve as discussants are invited to submit their name, address, professional affiliation, position, telephone number, and major areas of interest *within* the history of economics. All communications regarding the conference should be addressed to Professor George J. Stigler, Haskell Hall, University of Chicago, Chicago, Illinois 60637.

On May 20, 1976, the Economics Department of the City College of New York will sponsor an all-day conference on "The Economics of Information." The object of the conference is to stimulate research that is relevant to current public issues and policies and to economic analysis. Persons interested in reading papers at the conference are invited to submit them to Professor Robert D. Leiter or Professor Malcolm Galatin, Department of Economics, City College of the City University of New York, Convent Avenue and West 138 Street, N.Y. 10031. Papers delivered at the conference are scheduled for publication in the fourth annual conference volume of the department.

The meeting of the Gesellschaft für Wirtschafts-und Sozialwissenschaften (Verein für Socialpolitik), "Social Problems of Modern Industrial Societies," will be held September 13–15, September 1976 in Augsburg. The Society invites proposals for papers on: Structural unemployment; Pollution; Urbanization; The declining birthrate and the tendency towards small families; The systems of social security in relation to the social goals of redistribution and security, and the cost explosion in health services; Changes in the degree of unionization, its causes and its importance in union policy, and the strike as an instrument of union policy; The manpower approach to educational planning in developed economies, and the efficiency of the various systems of professional training schools; The conflict between redistribution and stabilization goals in fiscal policy, and the Social Budget as an instrument of public social policy. Applications are to be submitted to Professor Dr. B. Külp, Institut für Allgemeine Wirtschaftsforschung, D-7800 Freiburg i.Br., Werthmannplatz 1.

Visiting Economics Schclars Program

The Executive Committee of the American Economic Association has approved a Visiting Economics Scholars Program. The program has the following objectives: (1) to stimulate interest and help develop skills in economics among undergraduates in colleges and universities through short-term (one or two days) visits by distinguished members of the profession; (2) to provide opportunities for faculty members at the host institutions to discuss their teaching and research problems with the visitor. The host institution is expected to cover all travel and living expenses associated with the visit and to pay its usual honorarium. For more information, contact C. Elton Hinshaw, Assistant Secretary-Treasurer, American Economic Association, 1313 21st Avenue South, Nashville, Tennessee 37212.

Deaths

John C. Daniel, lecturer, department of economics, Rutgers College, Dec. 7, 1974.

Joseph S. Davis, professor emeritus of economic research, Food Research Institute, Stanford University, Apr. 23, 1975.

Leo Fishman, professor of economics, West Virginia University, Apr. 4, 1975.

Charles B. Friday, professor of economics and a former president of the Western Economic Association, Oregon State University, Apr. 16, 1975.

Calvin B. Hoover, Duke University.

George W. Stocking, June 7, 1975.

Retirements

Andrew J. Bartley, professor of economics, North Carolina State University, May 1975.

Frank T. de Vyver, Duke University.

Jean O. M. van der Mensbrugghe, advisor, International Monetary Fund, Mar. 31, 1975.

James S. Youtsler, professor and chairman, department of economics, Skidmore College, July 1975.

Visiting Foreign Scholars

Robin Marris, King's College, Cambridge University: visiting professor of economics, University of Maryland, 1975–76.

Ezra J. Mishan, London School of Economics: visiting professor of economics, University of Maryland, fall semester 1975.

David K. Round: visiting lecturer, department of economics, Duke University.

Promotions

George J. Bossy: manager, foreign department, Federal Reserve Bank of New York, Feb. 21, 1975.

Robert M. Fearn: professor of economics and business, North Carolina State University, July 1975.

Joseph W. Ford: professor of economics, St. Francis College, Sept. 1, 1975.

A. Ronald Gallant: associate professor of statistics and economics, North Carolina State University, July 1975.

Daniel A. Graham: associate professor, department of economics, Duke University.

Christopher I. Higgins: assistant secretary, economic Branch, Australian Treasury, Apr. 1975.

George R. Juncker: chief, Bank Analysis Division, Federal Reserve Bank of New York, Jan. 9, 1975.

Hugh L. Liner: extension professor of economics and business, North Carolina State University, July 1975.

Jerome L. McElroy: associate professor of economics, College of the Virgin Islands, Sept. 1975.

Carlisle E. Moody: associate professor, College of William and Mary, Sept. 1975.

Richard W. Nelson: chief, Banking Studies Division, Federal Reserve Bank of New York, Jan. 9, 1975.

Clayne L. Pope: associate professor of economics, Brigham Young University, Sept. 1975.

Hiroki Tsurumi: professor of economics, department of economics, Rutgers College, July 1, 1975.

Joan G. Walters: professor of economics, Fairfield University, Sept. 1974.

Administrative Appointments

Roger E. Bolton: acting chairman, department of economics, Williams College, Jan. 1975.

Thomas M. Freeman: director, Office of Institutional Research, Michigan State University, Mar. 1, 1975.

Martin A. Garrett: chairman, department of economics, College of William and Mary, Feb. 1, 1975.

George J. Neimanis: chairman, department of economics and commerce, Niagara University, June 1, 1975.

Clayne L. Pope: chairman, department of economics, Brigham Young University, Sept. 1975.

Kenan Sahin: associate chairman, department of management science and management information systems, University of Massachusetts.

Sidney C. Sufrin: chairman, department of general business and finance, School of Business Administration, University of Massachusetts.

Lawrence A. Weiser: chairman, department of economics and business administration, University of Wisconsin, Stevens Point, Aug. 25, 1975.

New Appointments

Joseph M. Anderson, Harvard University: assistant professor of economics, Williams College, July 1, 1975.

Kenneth C. Baseman, Stanford University: staff economist, Economic Policy Office, Antitrust Division, U.S. Department of Justice.

Douglas H. Blair: lecturer, department of economics, Rutgers College, July 1, 1975.

Bruce Bolnick: assistant professor, department of economics, Duke University.

Yale Braunstein: assistant professor, department of economics, New York University, Sept. 1974.

James M. Broadus, Yale University: staff economist, Economic Policy Office, Antitrust Division, U.S. Department of Justice.

Charles C. Brown, Harvard University: assistant professor of economics, University of Maryland, Aug. 15, 1975.

Marie E. Brown, University of California, Los Angeles: assistant economist, economics department, The Rand Corporation, Dec. 1974.

Anne J. Bynoe: lecturer, department of economics, Rutgers College, July 1, 1975.

Robert L. Clark: assistant professor of economics and business, North Carolina State University, fall 1975.

John F. Cogan, University of California, Santa Barbara: associate economist, economics department, The Rand Corporation, July 1975.

Neal P. Cohen: assistant professor of economics, St. Louis University, Sept. 1, 1975.

Eugene P. Durbin, Technology Service Corporation: senior economist, economics department, The Rand Corporation, Jan. 1975.

Philip M. Eisenstat, Northwestern University: staff economist, Economic Policy Office, Antitrust Division, U.S. Department of Justice.

Norman Familant, University of Virginia: staff economist, Economic Policy Office, Antitrust Division, U.S. Department of Justice.

Richard S. Fenwick, Jr., Kansas State University: associate professor of economics and business, North Carolina State University, July 16, 1975.

Gerald L. Fralick, U. S. Postal Rate Commission: staff economist, Economic Policy Office, Antitrust Division, U.S. Department of Justice.

John F. Geweke: assistant professor, department of economics, University of Wisconsin, Madison.

Marion H. Gillim: professor of economics, Eastern Kentucky University, Aug. 1974.

Peter T. Gottschalk, University of California, Santa Cruz: assistant professor of economics, Mount Holyoke College, fall 1975.

Frederick W. Gramlich, Harvard University: staff economist, Economics Policy Office, Antitrust Division, U.S. Department of Justice.

Frederick H. deB. Harris: assistant professor of economics, College of William and Mary, Sept. 1975.

Donald R. Haurin, University of Chicago: assistant professor, Ohio State University.

Ronald A. Heiner: assistant professor of economics, Brigham Young University, Sept. 1975.

Peter H. Henning, University of Michigan: assistant professor, Wayne State University, Sept. 1975.

Susan Hosek: associate economist, economics department, The Rand Corporation, July 1975.

John T. Hoven, University of Wisconsin: staff economist, Economic Policy Office, Antitrust Division, U.S. Department of Justice.

Joseph P. Hughes: assistant professor, department of economics, Rutgers College, July 1, 1975.

James B. Kearl: assistant professor of economics, Brigham Young University, Sept. 1975.

Kenneth J. Kopecky, Ohio State University: economist, Federal Reserve Board, Sept. 1975.

Arthur Kraft: professor of management science, University of Nebraska, Lincoln, Aug. 1975.

Martha J. Langer, Yale University: department of economics, Mount Holyoke College, fall 1975.

Augusto A. Legasto, Jr.: assistant professor, department of economics, Rutgers College, July 1, 1975.

Edgar W. Leonard: assistant professor of economics and business, North Carolina State University, fall 1975.

Charles Lieberman, University of Pennsylvania: assistant professor of economics, University of Maryland, Aug. 1974.

Martin P. Loeb: assistant professor of economics and business, North Carolina State University, fall 1975.

Jerome L. McElroy, College of the Virgin Islands: senior research associate and socioeconomic research coordinator, Caribbean Research Institute, Sept. 1974.

J. Michael McGuire: assistant professor of economics, St. Louis University, Sept. 1, 1975.

Willard G. Manning, Jr., Harvard University: associate economist, economics department, The Rand Corporation, July 1975.

Neil W. Mather, University of Wisconsin: assistant professor, Wayne State University, Sept. 1975.

Edgar O. Olsen: visiting associate professor of economics, University of Wisconsin, Madison; project associate, Institute for Research on Poverty.

A. Mead Over, Jr., University of Wisconsin: assistant professor of economics, Williams College, July 1, 1975.

John D. Owen, New School for Social Research: professor, Wayne State University.

Wayne D. Perry, Carnegie-Mellon University: associate economist, economics department, The Rand Corporation, June 1975.

Andrew Popper: economist, Foreign Research Division, Federal Reserve Bank of New York, Mar. 20, 1975.

Arturo C. Porzecanski, University of Pittsburgh: economist, Centro de Estudios, Monetarios Latinoamericanos, Mexico City, Sept. 1, 1975.

David L. Roberts: assistant professor, department of economics, Rutgers College, July 1, 1975.

Samuel Rosenberg, University of California, Davis: assistant professor of economics, Williams College, July 1, 1975.

Clark G. Ross: assistant professor of economics, College of William and Mary, Sept. 1975.

Kevin T. Saiki: instructor, economics, U.S. International University, Makawao, Hawaii.

David A. Schauer: assistant professor of economics and finance, University of Texas, El Paso, Sept. 1, 1975.

Jeff A. Schnepper: lecturer, department of economics, Rutgers College, July 1, 1975.

Peter Simon, University of Pennsylvania: staff economist, Economic Policy Office, Antitrust Division, U.S. Department of Justice.

Lee I. Sparling, California Institute of Technology: staff economist, Economic Policy Office, Antitrust Division, U.S. Department of Justice.

Gordon R. Storholm: lecturer, department of economics, Rutgers College, July 1, 1975.

Frederick R. Strobel, Federal Bank of Atlanta: associate professor of money and banking, Kalamazoo College, Jan. 1, 1975.

Miguel L. Tejada, Iowa State University: senior specialist, Unit of Technical Change and Transfer of Technology, Organization of American States, Washington, Sept. 1, 1974.

Edward Tower: associate professor, department of economics, Duke University.

T. Dudley Wallace: professor, department of economics, Duke University.

Geraldine A. Walter, University of California, Los Angeles: assistant economist, economics department, The Rand Corporation, Mar. 1975.

Bernard L. Weinstein, State University of New York: senior assistant professor, department of economics, University of Texas, Dallas, Sept. 1975.

Walter J. Wessels: assistant professor of economics and business, North Carolina State University, fall 1975.

Robert C. West, Northwestern University: assistant professor of economics, University of Maryland, Aug. 1974.

Robert W. Wilson, Yale University: staff economist, Economic Policy Office, Antitrust Division, U.S. Department of Justice.

Allen M. Wyse: assistant professor, department of economics, Duke University.

Leaves for Special Appointments

Robert P. Armstrong, International Bank for Reconstruction and Development: economic advisor to the Ministry of Economy and Economic Cooperation, Ford Foundation Investment Planning Project, Arab Republic of Egypt, Feb. 1975.

Leslie E. Grayson, University of Virginia: Ford Foundation award, visiting professor, European Graduate Institute of Business Administration, 1975–76.

Karl D. Gregory, Oakland University: senior staff member, Congressional Budget Office, U.S. Congress, Washington, June 1975.

Ira Horowitz, University of Florida: staff economist, Economic Policy Office, Antitrust Division, U.S. Department of Justice.

William Jaffé, York University: Killam senior research scholarship, 1975–76.

Richard K. Perrin, North Carolina State University: associate economist, CIMMYT, Mexico, Jan. 1, 1975.

Leonard G. Schifrin, College of William and Mary: National Science Fellow, School of Medicine, Stanford University, 1975–76.

Norman P. Swenson, Corps of Engineers: planning associate, Board of Engineers for Rivers and Harbors, Aug. 11, 1974.

Richard E. Sylla, North Carolina State University: senior associate, St. Antony's College, Oxford University, Sept. 1975–June 1976.

Resignations

Ashok Bhargava, University of Wisconsin, Whitewater, May 1975.

Leon M. Ennis, Jr., North Carolina State University: University of North Carolina, Chapel Hill, July 1975.

Thomas O. McCoy, Williams College, June 30, 1975.

Robert C. Rice, University of Hawaii: Monash University, Australia, Aug. 1975.

Henry Thomassen, University of Nebraska, Lincoln.

Rodger Weaver, University of Nebraska, Lincoln, Jan. 5, 1975.

NOTE TO DEPARTMENTAL SECRETARIES AND EXECUTIVE OFFICERS

When sending information to the *Review* for inclusion in the Notes Section, please use the following style:

A. Please use the following categories:

1—Deaths
2—Retirements
3—Foreign Scholars (visiting the USA or Canada)
4—Promotions
5—Administrative Appointments
6—New Appointments
7—Leaves for Special Appointments (NOT Sabbaticals)
8—Resignations
9—Miscellaneous

B. Please give the name of the individual (SMITH, John W.), his present place of employment or enrollment: his new title (if any), his next place of employment (if known or if changed), and the date at which the change will occur.

C. Type each item on a separate 3×5 card, and please do not send public relations releases.

D. The closing dates for each issue are as follows: *March*, November 1; *June*, February 1; *September*, May 1; *December*, August 1.

This announcement supersedes and replaces a letter which was sent annually from the managing editor's office. All items and information should be sent to the Assistant Editor, *American Economic Review*, Box Q, Brown University, Providence, Rhode Island 02912.

◆

DETENTE AND DOLLARS:

Doing Business with the Soviets

Marshall I. Goldman

This indispensable book by a leading expert in East-West trade, provides the American businessman with the perspective and information he needs to decide whether he too should enter the billion-dollar Soviet market and, if so, how he can do it successfully. Dr. Goldman draws important lessons from his detailed inside account of both costly failures (the great American grain disaster of 1972) and success stories (Armand Hammer and the Pullman Corporation) in doing business with the Soviet Union.
Due June $10.00

HOUSING AND THE MONEY MARKETS

Roger Starr

Here is the first systematic introduction to the money markets as they relate to housing. Now Administrator of New York City's Housing and Development Administration, Mr. Starr describes the various institutions that extend mortgages and shows how they finance housing construction. He explores the relationship between lending institutions and the various government agencies such as the FHA, GINNY MAE, and FANNY MAE that have been set up for channeling savings into mortgages. "Timely as well as valuable.... An excellent primer and a well reasoned policy text for specialists and government officials."—LOUIS WINNICK.
Just published $10.95

ANARCHY, STATE AND UTOPIA

Robert Nozick

In the most challenging defense of property rights and free enterprise in modern times, Professor Nozick poses, as *Forbes* puts it, "nothing less than a challenge to a central feature of the moral landscape: the state as everybody's mother and father."—*Forbes*. "No contemporary philosopher possesses a more imaginative mind, broader interests, or greater dialectic abilities."—*Harpers*. Published $12.95

WHO SHALL LIVE?

Health, Economics, and Social Choice

Victor R. Fuchs

"A short lucid, uncomplicated explanation of the economics of health care.... No book has demonstrated a better grasp of the *social* context of health and health care, of the idea that central social changes, rather than mere technical tinkering with the system is at issue. Most useful."
—*The New York Times Book Review*.
Just published $8.95

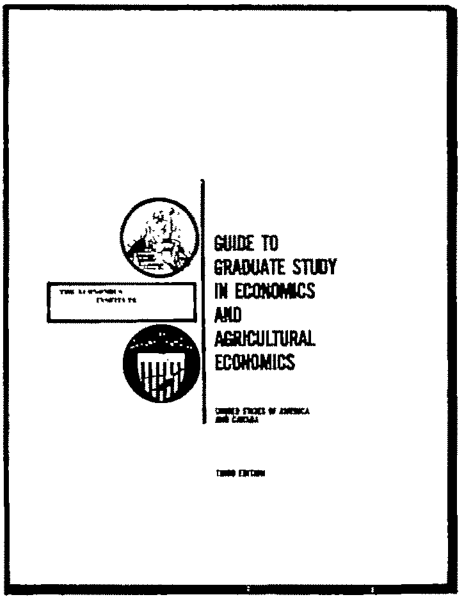
CRISIS IN KEYNESIAN ECONOMICS

Sir John Hicks

Just published $5.95

NEW. SIGNIFICANT.
BASIC
BASIC BOOKS INC.
10 EAST 53RD ST, NEW YORK 10022

Announcing...



GUIDE TO GRADUATE STUDY IN ECONOMICS AND AGRICULTURAL ECONOMICS

Guide to Graduate Study in Economics and Agricultural Economics

in the United States of America and Canada

THIRD EDITION

Designed to provide students anticipating graduate study in economics and agricultural economics, and their advisors, with information on available graduate training programs.

Includes descriptions of 236 graduate programs, supplemented by comparative data and information for prospective students, domestic and foreign.

Published by the Economics Institute under the auspices of the American Economic Association, and the American Agricultural Economics Association.

Distributed by:
RICHARD D. IRWIN INC.,
1818 Ridge Road,
Homewood, Illinois 60430

PRICE: $7.50 per copy

"The topics are timely... the materials are pertinent... a welcome contribution."

—Michael Yoshino, Professor of Business Administration, Harvard University.

This is the first in-depth analysis of Japanese business ever written in English.

The authors, both specialists in the field, tell the success story of Japanese business in general and of firms like Sony, Honda, and Mitsubishi in particular. And they explain how the Japanese have managed to incorporate traditional values into industrial empires.

Essential reading for anyone doing business on an international level. Or studying it. $12.00

The Development of Japanese Business, 1600-1973

by Johannes Hirschmeier
and Tsunehiko Yui

Harvard
University Press
Cambridge, Mass. 02138

Begin making plans to attend the

Eighty-eighth
Annual Meeting of The American Economic Association

to be held in

DALLAS, TEXAS

Sunday, Monday, Tuesday

DECEMBER 28-30, 1975

The Employment Center opens Saturday, December 27.
This is a change from the October dates previously announced.

See the Notes section of the September, AER for the preliminary program.

The 1976 meeting will be held in Atlantic City, September 16-18.

PRAEGER
SPECIAL STUDIES

THE UNITED STATES AND
INTERNATIONAL OIL
A Report for the Federal Energy
Administration on U.S. Firms and
Government Policy
Robert B. Krueger
772 pp. (incl. appendixes) May, 1975
ISBN 0-275-05300-8 $45.00
PSS Student Edition: 384 pp. (no appendixes)
ISBN 0-275-89340-5 $8.50

ECONOMIC ANALYSIS AND THE
MULTINATIONAL ENTERPRISE
Edited by John H. Dunning
405 pp. June, 1975 ISBN 0-275-09900-8 $20.00

THE UNITED STATES AND
WORLD DEVELOPMENT
Agenda for Action, 1975
James W. Howe and the Staff of the
Overseas Development Council
288 pp. June, 1975 ISBN 0-275-05260-5 $16.50
PSS Student Edition: ISBN 0-275-89310-3 $4.50

SCIENCE POLICIES OF
INDUSTRIAL NATIONS
Case Studies of the United States, Soviet
Union, United Kingdom, France,
Japan, and Sweden
Edited by T. Dixon Long and
Christopher Wright
250 pp. June, 1975 ISBN 0-275-05600-7 $17.50

SOVIET INDUSTRIAL
IMPORT PRIORITIES
With Marketing Considerations for
Exporting to the USSR
Christopher E. Stowell, assisted by Neal
Weigel, with chapters by Erast Borissoff
and Edward Maguire
500 pp. June, 1975 ISBN 0-275-09490-1 $30.00

THE SCOPE OF BARGAINING
IN PUBLIC EMPLOYMENT
Joan Weitzman
404 pp. June, 1975 ISBN 0-275-05760-7 $23.50

REGIONAL ECONOMIC GROWTH
Diversification and Control
Michael E. Conroy
186 pp. July, 1975 ISBN 0-275-05180-3 $15.00

THE NATION-STATE AND
TRANSNATIONAL CORPORATIONS
IN CONFLICT
With Special Reference to Latin America
Edited by Jon P. Gunnemann
272 pp. July, 1975 ISBN 0-275-05200-1 $17.50

PERSONAL AND SOCIAL
CONSUMPTION IN EASTERN EUROPE
Poland, Czechoslovakia, Hungary, and
East Germany
Bogdan Mieczkowski
368 pp. July, 1975 ISBN 0-275-05620-1 $21.50

ECONOMIC INTEGRATION AND
THE LAW OF GATT
Pierre Lortie
202 pp. Aug., 1975 ISBN 0-275-05230-3 $16.50

TAXES ON DIRECT INVESTMENT
INCOME IN THE EEC
A Legal and Economic Analysis
Bernard Snoy
382 pp. Aug., 1975 ISBN 0-275-05710-0 $21.50

THE SAUDI ARABIAN ECONOMY
Ramon Knauerhase
400 pp. Sept., 1975 ISBN 0-275-09000-0 $23.50

THE STATES AND
LAND-USE CONTROL
R. Robert Linowes and Don T. Allensworth
200 pp. Sept., 1975 ISBN 0-275-05210-9 $17.50

INDUSTRIAL LOCATION DECISIONS
Detroit Compared with Atlanta and Chicago
Lewis Mandell
112 pp. Sept., 1975 ISBN 0-275-05610-4 $12.50

Please use ISBN or TC numbers when
ordering from:

praeger publishers
111 Fourth Avenue
New York, New York 10003

Nicolaus & Associates, Inc.

has been appointed
official travel coordinator
for the meetings of the
Allied Social Science Associations
in Dallas, Texas
December 28-30, 1975

The Employment Center opens Saturday, December 27.

Group air fares may be available from your city and for groups as small as fifteen. These fares usually run about 25% below regular coach fares. For example current fares from three cities:

	Coach	*Group*
New York-Dallas-New York	236.73	159.73
Des Moines-Dallas-Des Moines	124.73	99.73
San Francisco-Dallas-San Francisco	252.73	202.73

Air reservation forms and general information on flights and costs will be mailed to you in October with the pre-registration material. In the meantime if you have any questions on group flights, please call Nicolaus & Associates collect at area code 817-461-1622 or write them at 1201 N. Watson Road, Suite 179,
Arlington, Texas 76011.

BLUE CROSS SINCE 1929:
Accountability and The Public Trust

Odin W. Anderson, Center for Health Administration Studies, University of Chicago

An analysis of the relationship between government and the private sector in providing health services.
128 Pages $10.00

THE REGULATORY AGENCY AND THE PUBLIC CORPORATION:
The Canadian Air Transport Industry

John R. Baldwin, Queens University

How government agencies can influence industry performance positively and for the public good.
170 Pages $12.50

THE POLITICS OF SOCIAL PROGRAM EVALUATION

David K. Banner, University of New Brunswick, Samuel I. Doctors, University of Pittsburgh, and Andrew C. Gordon, Northwestern University

Seeks an objective method of scientific proof of the efficacy of government funded social action programs.
224 Pages $13.50

REGULATING THE PRODUCT:
Quality and Variety

Richard E. Caves and Marc J. Roberts, Harvard University, Editors

Specific case studies demonstrate how regulatory influences affect the quality and variety of products.
256 Pages $13.50

THE ECONOMICS OF PROPERTY RIGHTS

Eirik G. Furubotn, Texas A & M University and Svetozar Pejovich, Ohio University, Editors

Papers which emphasize the interconnectedness of ownership rights, incentives, and economic behavior.
398 Pages (cloth) $19.50 (paper) $7.95

THE NEW POLITICAL ECONOMY OF THE PACIFIC

Bernard K. Gordon and Kenneth J. Rothwell, University of New Hampshire, Editors

Essays considering how political and economic changes in the Pacific Basin will affect the world system. 192 Pages $11.50

ENERGY POLICY – INDUSTRY PERSPECTIVES

John E. Gray (Energy Policy Project of the Ford Foundation)

The views of energy industry leaders on energy problems and goals – government policy, laws and regulations. 144 Pages (cloth) $8.50 (paper) $2.95

INDUSTRYWIDE VOLUNTARY PRODUCT STANDARDS

David Hemenway

The effect of market structure on both incentive and ability of industry to create workable product standards. 160 Pages $12.50

PUBLIC NEEDS AND PRIVATE BEHAVIOR IN METROPOLITAN AREAS

John E. Jackson, Harvard University, Editor

Alternatives and implications of the ways by which metropolitan residents satisfy their demands for public services. 184 Pages $13.50

WORLD ENERGY STRATEGIES:
Facts, Issues and Options

Amory B. Lovins (Co-Published with Friends of the Earth International)

Detailed explanation of technical elements in disputes about our inadequate energy resources.
166 Pages (cloth) $10.95 (paper) $4.95

FARMERS' COOPERATIVES AND FEDERAL INCOME TAXES

Lee F. Schrader, Purdue University and Ray A. Goldberg, Harvard University

Management strategies for farm cooperatives based on 1973 income-tax policies. 120 Pages $12.50

TRADE, AGRICULTURE, AND DEVELOPMENT

George S. Tolley, University of Chicago, Editor

The roles of agriculture and trade in development – theories for quickening growth and empirical analyses of several developing countries. 224 Pages $13.50

THE ENERGY CONSERVATION PAPERS

Robert Williams, Editor (Energy Policy Project of the Ford Foundation)

Fresh facts and new insights into energy conservation.
416 Pages (cloth) $17.50 (paper) $8.95

For information on these and other Ballinger titles, please write for our catalog.

Ballinger

17 Dunster Street
Harvard Square, Cambridge, MA 02138

The ECONOMICS INSTITUTE

SPECIAL PROPAEDEUTIC TRAINING AND ORIENTATION PROGRAM FOI

INTERNATIONAL

Graduate Students

in ECONOMICS and RELATED FIELDS

UNIVERSITY OF COLORADO, JANUARY 19 - AUGUST 18, 1976

With six consecutive sessions beginning January 19, February 23, March 29, May 3, June 7 and July 12, 1976.

The Institute is specially designed for students from other countries who are planning to undertake graduate studies at universities in the United States in economics and related fields, including business and agricultural economics. An intensive and specialized *English* instructional program is progressively substituted by an expanding component of preparatory course offerings in *mathematics, statistics* and *economic theory,* and lectures and research seminars on special field topics including agriculture, business and development. The program also includes field trips and social recreational activities designed to acquaint the participants with campus and community life in the United States.

Admissions to the sessions beginning in January, February, March, and May are restricted to students needing further intensive training in English; those who would score in excess of 500 on the TOEFL at the point of registration should apply for the later (summer) sessions.

Universities and sponsoring agencies are invited to utilize the Institute's Credentials Evaluation Service and its Spring and Summer testing and training programs in their evaluation of candidates for graduate program admissions and related financial support.

Further information and application materials may be obtained from the DIRECTOR, ECONOMICS INSTITUTE, UNIVERSITY OF COLORADO, BOULDER, COLORADO 80302, U.S.A.

Program sponsored by th

American Economic Associatio

in cooperation wit
The American Assembly of Collegiate Schools of Busines
The American Agricultural Economics Associatio
The Institute of International Educatio
and THE UNIVERSITY OF COLORAD

Law School of Harvard University
Cambridge, Mass. 02138

Liberal Arts Fellowships in Law

For the academic year 1976–1977 Harvard Law School offers four or five Liberal Arts Fellowships to college and university teachers in the arts and sciences for a year at the Law School. Holders of these Fellowships will have the title of Fellow in Law and . . . (History, Sociology, Political Science, Economics, Philosophy, etc., depending upon their particular discipline).

The purpose of the fellowships is to enable teachers in the social sciences or humanities to study fundamental techniques, concepts, and aims of law, so that, in their teaching and research, they will be better able to use legal materials and legal insights which are relevant to their own disciplines.

Fellowship holders will presumably take at least two first-year courses in law, in addition to more advanced courses, and will participate in a joint seminar. The year of study will not count toward a degree.

The fellowship grant is sufficient to cover tuition and health fees. The Chairman of the Liberal Arts Fellowship Committee will be glad to write a letter to any funding agency to which the applicant has applied describing the Program and indicating the extent of the Committee's interest in inviting the applicant to be a Fellow.

Applications should include a biographical résumé (including academic record and list of publications), a statement explaining what the applicant hopes to achieve through his year of study, and two letters of recommendation to be sent directly by the referees.

Applications for 1976–1977 should be submitted before January 15, 1976, to the Chairman, Committee on Liberal Arts Fellowships, Harvard Law School, Cambridge, Massachusetts 02138.

Awards will be announced before February 15, 1976.

Wiley/Hamilton: Business, but not as usual Hamilton Publishing Company has a new point of view: *yours.* Under the new imprint of Wiley/Hamilton, the four-year-old subsidiary of John Wiley & Sons has assumed responsibility for all college-level publishing activities related to the fields of business and economics.

Consolidation of all business-related textbook publishing into a single publishing unit has made it possible for Wiley to assemble the specialized editorial, marketing, and production talents of three corporate groups: Hamilton, the Wiley College Division, and Melville Publishing Company. Narrowing and deepening their publishing program will enable the Wiley/Hamilton editorial, marketing, and production specialists to direct their undivided attention to your areas of interest. The new Wiley/Hamilton group will offer textbook authors in the economics and business-related disciplines the kind of close personal attention that has been a part of Hamilton's philosophy from the company's inception.

If you have textbook needs, or plan to develop a textbook in your area of interest, consider a company that shares your point of view.

Let Wiley/Hamilton mind *your Business.*

Available:

MACROECONOMICS:

Measurement, Theory, and Policy
By Michael C. Lovell,
Wesleyan University

This is an outstanding text for intermediate and upper-level courses in macroeconomics. Lovell stresses modern theoretical advances in analyzing unemployment, inflation, balance of payments, disequilibrium, business cycles, growth.

Schematic flowcharts show students how the equations which constitute the building blocks of a sequence of models fit together. Each chapter ends with a list of key concepts, exercises, and problems.

1975 $14.95

ECONOMICS AND SOCIETY

By Leonard W. Weiss,
University of Wisconsin, Madison

Some economics books stand out. Among these few, Weiss is simply outstanding.

ECONOMICS AND SOCIETY builds its analysis around supply and demand, and aggregate supply and demand. But the theory remains as straightforward as possible—never presented for its own sake. Rather, concepts are applied to such policy issues as monopoly, discrimination, labor market, balance of payments, and public goods. While kept elementary, every concept is developed in the context of realistic, practical experience...to give your students the major elements of economics in a meaningful, realistic setting.

1975 $10.95

Forthcoming:

MICROECONOMICS ANALYSIS WITH APPLICATIONS

By Robert Y. Awh,
Mississippi State University

Awh offers a balanced, comprehensive coverage of up-to-date intermediate microeconomics, combining a solid treatment of theory with real-world applications.

1976

MACROECONOMICS:

Money, Prices, and Income
By Dudley W. Johnson,
University of Washington

This book presents an integrated approach to macroeconomics; that is, it covers the traditional subject matter, as well as the economics of money and banking, and international monetary economics. It also provides an exhaustive analysis of monetarism versus fiscalism.

November 1975 $11.95 (tent.)

MACROECONOMICS Third Edition

By John Lindauer,
Murray State University

This is a thorough, straightforward, and graphic treatment of macroeconomics. This edition has been revised to provide an up-to-date analysis of the state of the art.

February 1976 $13.95 (tent.)

INTRODUCTION TO ECONOMICS:

An Issues Approach

By Chris Rhoden,
Solano College

Pedagogically sound, extremely readable: Rhoden's INTRODUCTION TO ECONOMICS features objectives, summaries, glossaries, and discussion questions for each chapter—as well as related readings drawn from Business Week, Wall Street Journal, etc. Also included: brain teasers (projects for student research).

This introduction to basic concepts and contemporary issues is perfect for the one-term basic economics course. Available with an Instructor's Manual and a Student Study Guide.

January 1976 $9.95 (tent.)

For your complimentary examination copies, contact your Wiley representative or write to George Thomsen, Dept. 840, at the address below. Please include course title, enrollment and present text.

WILEY/HAMILTON
The Business and Economics
Publishing Group of Wiley,
1129 State Street,
Santa Barbara, California 93101

Prices subject to change without notice. A5166-HP

NEW
FROM CALIFORNIA

The Economic Development of the Third World since 1900

Paul Bairoch
Translated by Lady Cynthia Postan

First published in 1967, Professor Bairoch's *Diagnostic de l'évolution économique du Tiers-Monde* has gone into five editions, and has brought the author an international reputation.

274 pages, 53 tables, $12.00

At bookstores

1900 1920 1940 1960 1980

UNIVERSITY OF CALIFORNIA PRESS
BERKELEY 94720

New Penguins in ECONOMICS

SELF-MANAGEMENT: Economic Liberation of Man. *Edited and introduced by Jaroslav Vanek.* Readings on the doctrine of self-management, actual cases, performance, and economic theory. $4.95(t)

MARXIAN AND POST-MARXIAN POLITICAL ECONOMY: An Introduction. *Arun Bose.* $5.95

AN ECONOMIC HISTORY OF THE USSR (Reissue). *Alec Nove.* $4.95

RICH AGAINST POOR. *C.R. Hensman.* The real implications of foreign aid. $3.95

THE MEDIEVAL ECONOMY AND SOCIETY. *M.M. Postan.* Completes *The Pelican Economic History of Britain* with *Reformation to Industrial Revolution* and *Industry and Empire*. $3.50

From the backlist two standard adoption texts:

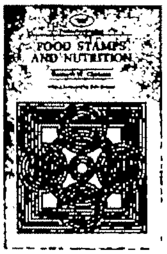
A CRITIQUE OF ECONOMIC THEORY. *Edited by E.K. Hunt and Jesse G. Schwartz.* $4.75

ECONOMIC SYSTEMS AND SOCIETY: Capitalism, Communism and the Third World. *George Dalton.* $3.95

For a complete list of Penguin books in economics, write:

PENGUIN BOOKS INC
7110 Ambassador Road Baltimore, Md. 21207
Overseas readers write Penguin Education, Harmondsworth, Middlesex, England

Have Food Stamps Failed?



Food Stamps and Nutrition
by Kenneth W. Clarkson

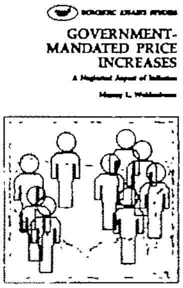
Food stamps do little to improve the nutrition of the poor and serve chiefly as an income subsidy to those who use them, Clarkson charges.

The study surveys Department of Agriculture statistics on the nutritional level of food stamp program participants, measuring the program's benefits to families of different income levels.

85 pages paperback, $3.00

Also from AEI:

Government-Mandated Price Increases
by Murray L. Weidenbaum



The author charges that federal regulatory actions are continually mandating higher prices. Examining such regulatory agencies as the Occupational Safety and Health Administration, Weidenbaum points out that each new benefit generates a new cost for the public. Although regulatory costs are less obvious causes of worsening inflation than budget deficits or "easy" monetary policies, the author says future regulatory actions must be carefully scrutinized to determine their actual costs. **112 pages paperback, $3.00**

Send for a complete catalog of AEI publications and audio and video cassettes.

AMERICAN ENTERPRISE INSTITUTE FOR PUBLIC POLICY RESEARCH
1150 Seventeenth Street, N.W., Washington, D.C. 20036
202/296-5616

OXFORD

Microeconomics

The Analysis of Prices and Markets

DONALD DEWEY, Columbia University. The lively and incisive writing of this text makes it one of the most readable introductions to modern microeconomic theory presently available. The text is the first to deal rigorously with the multiplant enterprise and market concentration. It contains an exceptionally detailed treatment of the theory of the firm, capital and interest, the pricing of factor inputs, and welfare economics. While the axiomatic character of modern microeconomic theory is firmly emphasized, its tools are applied to a wide range of policy problems that interest students: conservation of natural resources, regulations of public utilities, labor legislation, contracts of professional athletes, antitrust, and cost-benefit analysis of public projects.

1975 352 pp. 95 graphs $9.95

New books in the
ECONOMIC DEVELOPMENT SERIES
General Editor: GERALD M. MEIER,
Stanford University

Agriculture and Structural Transformation

Economic Strategies in Late-Developing Countries

BRUCE F. JOHNSTON, Stanford University, and **PETER KILBY,** Wesleyan University. In this volume, an agricultural and an industrial economist examine the problems of designing agricultural strategies that will help overcome the acute poverty that characterizes low-income nations. They analyze the interrelationships between agriculture and other sectors of the economy and consider the special problems and opportunities associated with late development. "This is an excellent study of the policy alternatives for late-developing countries in regard to food and agriculture. It appraises the general problem and suggests practical steps for food improvement."—Harold G. Halcrow, University of Illinois

1975 500 pp. graphs and tables cloth $11.95 paper $6.95

Employment, Technology, and Development

AMARTYA SEN, London School of Economics and Political Science. Focusing on the interrelationships among institutions, technology, and employment, the author investigates the institutional factors which affect employment policy making: modes of production, patterns of ownership, and systems of employment. The techniques of project evaluation and shadow pricing are discussed with particular reference not only to institutional features, but also to political feasibility and behavioral characteristics. Professor Sen illustrates his approach in four appendices which reflect India's experience in the field.

1975 200 pp. cloth $7.50 paper $3.50

All prices are subject to change.

Oxford Regional Economic Atlas of the United States and Canada

Second Edition

Advisory Editors: **JOHN D. CHAPMAN,** University of British Columbia, and **JOHN C. SHERMAN,** University of Washington. Prepared with the assistance of **QUENTIN H. STANFORD,** Don Mills Collegiate Institute, Toronto. Foreword by **JOHN H. PATERSON,** University of Leicester. Now in the Second Edition, *The Oxford Regional Economic Atlas of the United States and Canada* is a rich and exhaustive source of information on the resources, industries, and general trade of the two countries presented against a detailed picture of their physical environment. Every map retained from the first edition has been up-dated and several important new maps have been added.

Fall 1975 138 pages of maps gazetteer
cloth $21.95 paper $7.95

Structure of Classical Economic Theory

ROBERT B. EAGLY, State University of New York, Binghamton. "This book offers a clear, concise treatment of the essential structure and analysis of classical economics. It seems to me to be an excellent book . . . to bring together what students get in scattered form from reading several of the classical writers in selection." —Carl G. Uhr, University of California, Riverside. "This book is a significant contribution."—Daniel Coldwell, Memphis State University

1974 160 pp. $7.50

In the following trilogy of policy volumes in international political economy, Professor **GERALD MEIER** offers students a provocative approach to major policy problems involved in establishing a new international economic order. By blending specific source materials with a more general analytical narrative, the author immerses students in a set of actual policy-making situations so they can become concerned with the controversy of foreign economic policy analysis and discover criteria for judging the merits of alternative courses of action. "Excellent format! Students will be delighted. It is a unique way to present the important materials for the subject at hand." —Gerald M. Hampton, Lewis and Clark College

Problems of a World Monetary Order

1974 318 pp. paper $4.95

Problems of Cooperation for Development

1974 272 pp. paper $4.95

Problems of Trade Policy

1973 304 pp. paper $4.95

OXFORD UNIVERSITY PRESS
200 MADISON AVENUE
NEW YORK, N.Y. 10016

ESSAYS IN
NINETEENTH CENTURY
ECONOMIC HISTORY:

The Old Northwest (1975) $12.00

David C. Klingaman and Richard K. Vedder, Editors, Ohio University

These essays contributed by Donald R. Adams, Richard A. Easterlin, Lowell E. Gallaway, Robert E. Gallman, Don R. Leet, William N. Parker, Roger L. Ransom, Edward H. Rastatter, Lee Soltow, Jeffrey G. Williamson, Richard K. Vedder and David C. Klingaman focus on the factors and the institutions that played important roles in the economic growth of the United States during the nineteenth century. Covering a broad spectrum, the essays analyze the importance of agriculture, land, migration, wealth, banks, canals and the railroad to this growth.

"...the collection is of remarkably high quality and even when most innovative or controversial is always interesting to anyone concerned with the state, the region, or the nation in its nineteenth century experience." – MORTON ROTHSTEIN, University of Wisconsin.

OHIO UNIVERSITY PRESS
ATHENS, Ohio 45701

The 1975 LIFE INSURANCE FACT BOOK, reporting in detail the performance of the life insurance business during 1974, is available for distribution. THE FACT BOOK is a valuable aid and concise reference source for facts and figures on benefit payments, purchases, insurance in force, investments, income, state distributions and other information.

Single copies are free from the Institute of Life Insurance, Publications, Dept. E-2, 277 Park Avenue, New York, N.Y. 10017.

**1975
LIFE
INSURANCE
FACT
BOOK**

ACADEMIC PRESS

TEXTBOOKS IN ECONOMICS

Announcing:

MONEY AND INFLATION: A Monetarist Approach

J. HUSTON McCULLOCH, Boston College

At last, courses in money and banking can treat the monetarist theory of inflation at considerable depth. In this new supplementary textbook, Professor McCulloch examines both the theoretical and practical aspects of monetary theory from an unabashedly monetarist point of view.

- The treatment of the quantity theory goes far beyond the coverage available in most money and banking textbooks.
- The lively yet precise presentation brings together old and new extensions of the monetarist theory of inflation on a level the undergraduate can master without calculus.
- Problems and solutions reinforce the students' understanding of the critical concepts.

AVAILABLE FOR FALL CLASSES!
September, 1975, *about 120 pages, probable price $2.95*

Ready February, 1976

ECONOMICS: Private and Public Choice

JAMES D. GWARTNEY, Florida State University

An important new principles textbook.

Available Now:

INTRODUCTION TO URBAN ECONOMICS

DOUGLAS M. BROWN, Georgetown University

In this remarkably unified textbook, Professor Brown makes a detailed survey of the historical realities of urban life. Emphasis is placed on relating economic analysis to contemporary problems.

- The coverage of interurban and intraurban economic analysis requires only a background course in economic principles.
- The economic history of cities, from tribal organizations to the American city today, is surveyed. Particular attention is paid to colonial cities.
- Causes and possible solutions to critical urban problems are examined from the point of view of economic analysis with a minimal use of mathematics.

301 pages, $10.50

For complimentary copies, write to the Textbook Department.

ACADEMIC PRESS

A Subsidiary of Harcourt Brace Jovanovich, Publishers
111 FIFTH AVENUE, NEW YORK, N.Y. 10003

new books on **economic development**

The Assault on World Poverty

Problems of Rural Development, Education, and Health

With a preface by Robert S. McNamara, President, World Bank Group

Concerned with how to bring the benefits of development to the poor in developing countries by increasing their productivity, this book explores specific proposals for action in five related fields. *Published for the World Bank*

$17.50 hardcover, $5.95 paperback

Yugoslavia

Development with Decentralization
World Bank Country Economic Report

This report shows that the success with which Yugoslavia has combined rapid economic growth and fundamental changes over a long period of time speaks for the basic soundness of its mode of economic development. *Published for the World Bank*

$18.50 hardcover, $8.50 paperback

The Future for Hard Fibers and Competition from Synthetics

Enzo R. Grilli

Grilli examines the structure of the hard fibers economy, its problems and future prospects and includes detailed demand, supply, and trade forecasts to 1980 which provide a basic for adjustment determinations. *World Bank Staff Occasional Paper No. 19*

$5.00 paperback

A System of International Comparisons of Gross Product and Purchasing Power

Irving B. Kravis, Zoltan Kenessey, Alan Heston, and Robert Summers

The authors present both binary and multi-lateral comparisons of the gross domestic product per capita and currency purchasing power for ten countries, selected to reflect a variety of income levels, geographical locations, and economic systems. *Published for the World Bank*

$20.00 hardcover, $7.50 paperback

Readings on Taxation in Developing Countries

Third Edition

edited by Richard M. Bird and Oliver Oldman

This thoroughly revised, third edition describes not only the economic aspects of tax policies and systems, but also the legal and institutional factors of taxation in underdeveloped and developed countries. $16.50 hardcover

johns hopkins

The Johns Hopkins University Press
Baltimore, Maryland 21218

from Resources for the Future, Inc. new books on the economics of natural resources

U.S. Coal and the Electric Power Industry

Richard L. Gordon

Gordon examines current theories of coal economics and draws on extensive discussions with both coal consumers and those in the industry itself, to show that the future of coal as a fuel is not as promising as speculators suggest. $12.50 hardcover

The Economics of Natural Environments

Studies in the Valuation of
Commodity and Amenity Resources

John V. Krutilla and Anthony C. Fisher

Krutilla and Fisher present operational models for making quantitative estimates of the amenity value of environmental resources and illustrate with case studies which include the Hells Canyon dam and the Trans-Alaska pipeline.
$16.95 hardcover, $4.50 paperback

Conserving Energy

Prospects and Opportunities
in the New York Region

Joel Darmstadter

In this timely study, Darmstadter considers the full range of energy-conserving possibilities from political, social, and economic viewpoints.
$8.95 hardcover, $2.95 paperback

Water Management and Agricultural Development

A Case Study
of the Cuyo Region of Argentina

Kenneth D. Frederick

Frederick explores the emerging water scarcity in the arid Cuyo region — a region like many irrigated regions of the world — and discusses a range of policy options that would induce farmers to adopt more appropriate patterns of use. $10.00 hardcover

U.S. Energy Policy

Alternatives for Security

Douglas R. Bohi and Milton Russell

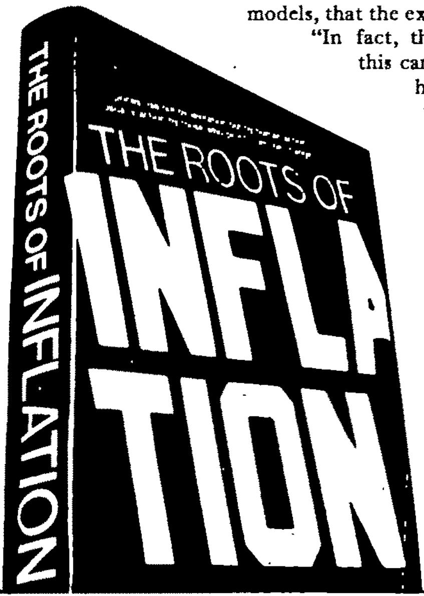
Responding to the search for energy security, Bohi and Russell propose a policy encouraging a limited quantity of imports, as well as energy and economic interdependence in order to reduce world oil prices.
$5.00 paperback

johns
hopkins

The Johns Hopkins University Press
Baltimore, Maryland 21218

"An extremely important and highly persuasive document..."

—JOHN KENNETH GALBRAITH



"It has now become fashionable to report that the economists have nothing to tell us, that a world of steeply rising prices and mounting unemployment fits none of the received models, that the experts owe us a decent silence....

"In fact, the intellectual state of economics is not as bankrupt as this caricature suggests. For nearly forty years, Gardiner Means has been mapping the outlines of a world not far removed from the one we now live in ... [and] John Blair has been carefully documenting what Means found....

"These are uncomfortable findings...."

—Bernard D. Nossiter (in *The Nation*)

"A valuable new book ... argues persuasively... explains the apparent paradox of simultaneous recession and inflation." —Robert Lekachman

"... a treasure house of ideas.."

—*The Yale Graduate Professional*

By GARDINER MEANS • JOHN BLAIR • *et al.*

Burt Franklin & Co., Inc. 235 E. 44, N.Y.C. 10017
Please send___copies of *Roots of Inflation* @$15.00
My check is enclosed. N.Y. residents add sales tax.
Name ________________
Address ________________
City, State, Zip ________________

STAFF PAPERS

Published by the International Monetary Fund three times a year with summaries in English, French, and Spanish.

Vol. XXII July 1975 No. 2

Floating Exchange Rates, Asymmetrical Intervention, and the Management of International Liquidity .... J. Marcus Fleming
Sharing the Oil Deficit .... Andrew D. Crockett and Duncan Ripley
Monetary and Exchange Rate Policies in a Small Open Economy .... Vicente Galbis
The Theory of Optimum Currency Areas: A Survey .... Yoshihide Ishiyama
International Aspects of the Taxation of Corporations and Shareholders .... Mitsuo Sato and Richard M. Bird
Dual Markets: The Case of the Syrian Arab Republic .... Paul M. Dickie and David B. Noursi
Some Empirical Evidence on the Determinants of Wage and Price Movements in Japan, 1950-73: A Survey .... Ichiro Otani
A Study of the Elasticity of the West Malaysian Income Tax System, 1961-70 ....Nurun N. Choudhry
Currency Arrangements and Banking Legislation in the Arabian Peninsula .... Michael E. Edo
Inflation Adjustment Schemes Under the Personal Income Tax .... Amalio Humberto Petrei
The Euro-Dollar Deposit Multiplier: A Note .... John Hewson and Eisuke Sakakibara

Staff Papers is one of four subscription periodicals published by the Fund: *Staff Papers* ($6 a year or $2.50 for a single copy), *Balance of Payments Yearbook* ($7.50), *Direction of Trade* ($10), and *International Financial Statistics* ($20). Special rates to university libraries, faculty members, and students: $3 a year for each of the first three titles (single copies of *Staff Papers* are $1) and $5 a year for *International Financial Statistics*; or all four periodicals for $12 a year. Subscriptions in other currencies accepted.

Address inquiries to The Secretary, International Monetary Fund, Washington, D.C. 20431, U.S.A.

CLASSICAL ECONOMICS RECONSIDERED

THOMAS SOWELL

The book evaluates classical economics in all its aspects, from its widely misunderstood social philosophy to its microeconomics, macroeconomics, and methodology. The author emphasizes classical economics as a system, treating particular doctrines and propositions as integral parts, rather than as isolated theories. Professor Sowell's reconsideration of the work of economists from Smith to Mill leads him to dispute many interpretations of classical economics in existing literature. $9.00

MULTINATIONAL CORPORATIONS AND THE POLITICS OF DEPENDENCE

Copper in Chile

THEODORE H. MORAN

Using the Chilean copper industry to study the relations between multinational corporations and their host countries in the Third World, the author describes the reaction against *dependencia*, a realization that the fate of the nation hinges on the decisions made by uncontrollable outside forces. *Written under the auspices of the Center for International Affairs, Harvard University* $12.50

CORPORATE GROWTH AND DIVERSIFICATION

CHARLES H. BERRY

As an increasing number of large corporations branch out into many fields of industry, public concern over the lateral extension of their power is aroused. Arguing that entry by large firms into concentrated industries may instead stimulate competition, Charles H. Berry analyzes the effect that such diversification has on corporate growth and on the structure and functioning of industrial markets. $9.50

Order from your bookstore or direct from
PRINCETON UNIVERSITY PRESS
Princeton, New Jersey 08540

Contributions to Economic Analysis

Editors:
J. TINBERGEN, D. W. JORGENSON, J. WAELBROECK

Volume 89:

EUROPEAN ECONOMIC INTEGRATION

edited by BELA BALASSA, Johns Hopkins University.

1975. about 430 pages. US$ 31.25/Dfl. 75.00. ISBN 0 7204 3193 X.

CONTENTS: The Common Market Experience (B. Balassa). Part I. Integration and Trade. The effects of European economic integration on the production and trade of manufactured products (E. M. Truman). An empirical examination of bilateral trade in Western Europe (S. A. Resnick and E. M. Truman). Trade creation and trade diversion in the European Common Market: An appraisal of the evidence (B. Balassa). Part II. Macroeconomic Implications of Integration. A macro-model for Western Europe (S. A. Resnick). Dynamic properties of the European macro-model (S. A. Resnick). Monetary integration in the European Common Market (B. Balassa). Part III. Policy Aspects of Integration. Structural policies in the European Common Market (B. Balassa). The effects of European economic integration on agriculture (E. Thorbecke and E. Pagoulatos). European integration and the developing countries (M. E. Kreinin). Appendices.

Volume 88:

ESSAYS ON ECONOMIC BEHAVIOR UNDER UNCERTAINTY

edited by M. BALCH, University of Iowa, D. McFADDEN, University of California at Berkeley, and S.WU, University of Iowa.

1974. 452 pages. US$ 51.95/Dfl. 125.00. ISBN 0 7204 3191 3.

The essays in this volume are selected from papers presented at the NSF-NBER Conferences on Decision Rules and Uncertainty. After an introductory essay, the volume is roughly divided into three parts. Part One comprises two papers that deal with the conceptual development of the conditional expected utility framework, Part Two includes five papers on various micro-aspects of behavior under uncertainty while the five papers in Part Three are concerned with welfare economics and general equilibrium.

Volume 87:

FRONTIERS OF QUANTITATIVE ECONOMICS, Vol. II

Papers presented at the Econometric Society Winter Meetings, Toronto, Canada, 1972

edited by MICHAEL D. INTRILIGATOR, University of California, Los Angeles, and DAVID A. KENDRICK, University of Texas at Austin.

1974. 592 pages. US$ 49.95/Dfl. 120.00. ISBN 0 7204 3190 5.

This volume forms a sequel to FRONTIERS OF QUANTITATIVE ECONOMICS – Papers Invited for Presentation at the Econometric Society Winter Meetings, New York, 1969 – edited by M. D. Intriligator (North-Holland, 1971).
The papers are grouped into three parts: The first part consists of papers on economic methodology and theory. The second is concerned with econometric techniques while the third covers quantitative approaches to traditional topics in economics.

Volume 86:

STUDIES IN BAYESIAN ECONOMETRICS AND STATISTICS

In Honor of Leonard J. Savage

edited by STEPHEN E. FIENBERG, University of Minnesota, and ARNOLD ZELLNER, University of Chicago.

1975. 686 pages. US$ 68.75/Dfl. 165.00. ISBN 0 7204 3188 3.

CONTENTS: Econometrics and Bayesian inference. Bayes' theorem and economic theory. Assessing prior distributions. Identification and Bayesian analysis. Applications of Bayesian regression. Missing observation problems. Time series problems. Analyses of simultaneous equation models. Control problems. Reporting results. New developments in estimation. Appendix. Index.

Volume 85:

POPULATION IN ECONOMIC GROWTH

by J. D. PITCHFORD, The Australian National University, Canberra.

1974. 288 pages. US$ 22.95/Dfl. 55.00. ISBN 0 7204 3187 5.

CONTENTS: Models of population growth. The determinants of net fertility. Population and economic growth. Optimum population, scale returns and trade. Optimum population and resources. Population in an optimal growth process. Resources and population control. Age-structure in population models. Conclusion. Index.

Volume 84:

INFLATION IN DEVELOPING COUNTRIES

An Economic Study of Chilean Inflation

by VITTORIO CORBO LIOI, Sir George Williams University, and International Institute of Quantitative Economics, Montreal, Canada.

1974. 300 pages. US$ 29.25/Dfl. 70.00. ISBN 0 7204 3186 7.

CONTENTS: Monetarism, Structuralism and Past Studies of Chilean Inflation. A Quarterly Model for Industrial Prices, Industrial Wages and Inflation. An Annual Study of Inflation. Appendices. References. Index.

north-holland
P.O. BOX 211
AMSTERDAM
THE NETHERLANDS

1567 NH

Distributed in the U.S.A. and Canada by
American Elsevier Publishing Company, Inc., 52 Vanderbilt Avenue, New York, N.Y. 10017

Studies in Mathematical and Managerial Economics

Edited by **H. THEIL**, Center for Mathematical Studies in Business and Economics, University of Chicago.

Volume 21: THE THEORY AND MEASUREMENT OF CONSUMER DEMAND, 2

by **H. THEIL**

1976. about 300 pages. In preparation.

CONTENTS: Conditional demand equations under block-independence. Blockwise dependence. Price-disturbance dependence. Statistical and geometric approaches to demand theory. The preference independence transformation. Changes in commodity baskets over time. Equivalent income changes, transitions between budget shares, and other topics. Bibliography. Index.

Volume 20: THE THEORY AND MEASUREMENT OF CONSUMER DEMAND, 1

by **H. THEIL**

1975. 350 pages. Clothbound ed.: US$ 15.95/Dfl. 38.00. ISBN 0 7204 3321 5
Paperback ed.: US$ 9.75/Dfl. 23.00. ISBN 0 7204 3095 X

CONTENTS: Utility theory. The Rotterdam Model. The indirect utility function and the economic theory of index numbers. Market demand. The estimation of complete systems of demand equations. The estimation of larger models. Bibliography. Index.

Volume 19: COST-BENEFIT ANALYSIS AND ECONOMIC THEORY

by **J. LESOURNE**, Chairman of the Board of Metra International.

1975. 530 pages. Clothbound ed.: US$ 41.25/Dfl. 99.00. ISBN 0 7204 3319 3
Paperback ed.: US$ 20.75/Dfl. 49.50. ISBN 0 7204 3097 6

CONTENTS: Introduction. Part I: Theory of Cost-Benefit Analysis. Some basic concepts. The analysis of marginal and structura transformations. The introduction of taxation. The comparison of growth programmes. The introduction of uncertainty as to the future. Information cost and decision cost. Part II: Applications of Cost Benefit Analysis. The industrial sector. The transport sector. Urban planning. Location of industry and the use of land. Public health. Education. General administration. Bibliography.

Volume 18: MANAGEMENT APPLICATIONS OF MODERN CONTROL THEORY

by **A. BENSOUSSAN, E. G. HURST** and **B. NÄSLUND.**

1974. 364 pages. Clothbound ed.: US$ 39.75/Dfl. 95.00. ISBN 0 7204 3318 5

CONTENTS: The need for dynamic management decisions. Introduction to modern control theory. Inventory and production control. Theory of the firm. Optimal maintenance and life of machines. Finance. Marketing. Control of natural resources. Numerical solutions of optimal control problems. Some limitations and further research. Mathematical review. Appendix A: Optimal deterministic control theory. Appendix B: Optimal stochastic control theory. References. Index.

Volume 17: METHODS FOR LINEAR AND QUADRATIC PROGRAMMING

by **C. VAN DE PANNE**, University of Calgary.

1975. 490 pages. Clothbound ed.: US$ 49.95/Dfl. 120.00. ISBN 0 7204 3316 9

CONTENTS: The solution of linear systems. The simplex method for linear programming. Variants of the simplex method. Duality in linear programming and the dual method. Sensitivity analysis and parametric programming. Parametric methods for linear programming. Post-optimality analysis, the reverse simplex method and related methods. Quadratic forms. The simplex and the dual method for quadratic programming. Parametric quadratic programming. Parametric methods for quadratic programming. Some other quadratic programming methods. Multiparametric linear and quadratic programming. Concave quadratic programming. General quadratic programming. The linear complementarity problem, I. The linear complementarity problem, II. Termination for copositive plus and £-matrices. Bibliography. Index.

Volume 16: CLASSICAL OPTIMIZATION: FOUNDATIONS AND EXTENSIONS

by **M. J. PANIK**, Austin Dunham Barney School of Business and Public Administration, University of Hartford, Connecticut, U.S.A.

1975. 260 pages. Clothbound ed.: US$ 26.95/Dfl. 65.00. ISBN 0 7204 3315 0
Paperback ed.: US$ 17.50/Dfl. 42.00. ISBN 0 7204 3081 X

CONTENTS: Mathematical foundations I. Mathematical foundations II. Global and local extrema of real-valued functions. Global extrema of real-valued functions. Polynomial approximation of real-valued functions. Local extrema of real-valued functions. Inflection points of real-valued functions. Convex and concave real-valued functions. Constrained optimization – a single equality side-relation. Constrained optimization – a generalization of the technique of Lagrange to *m* equality side-relations. Kuhn-Tucker theory – an introduction. Kuhn-Tucker theory – a generalization to *m* inequality side-relations. Kuhn-Tucker theory with mixed constraints. Bibliography.

Volume 15: ECONOMIC STABILIZATION POLICY – METHODS IN OPTIMIZATION

by **B. M. FRIEDMAN**, Assistant Professor of Economics, Harvard University.

1975. 394 pages. Clothbound ed.: US$ 33.50/Dfl. 80.00. ISBN 0 7204 3317 7

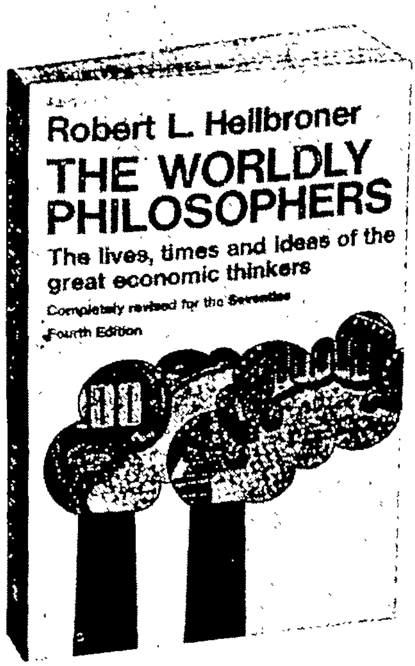
CONTENTS: Introduction. The econometric model. Point linearization. Subinterval linearization. The dynamic policy multipliers. Quadratic optimization – The level I problem. Piecewise quadratic optimization – The level II problem. Endogenous time horizon optimization – The level III problem. Applications to stabilization policy. Retrospective and prospective. Appendices A: List of Wharton model variable symbols. B: Linearized coefficient values for the Wharton model. C: Proof of piecewise quadratic convergence in the scalar case. D: The TRIOPT Algorithm. E: Card input guide for TRIOPT.

north-holland

P.O. BOX 211
AMSTERDAM
THE NETHERLANDS

1569 NH

Distributed in the U.S.A. and Canada by
American Elsevier Publishing Company, Inc., 52 Vanderbilt Avenue, New York, N.Y. 10017


Robert L. Heilbroner
THE WORLDLY PHILOSOPHERS
The lives, times and ideas of the great economic thinkers
Completely revised for the Seventies
Fourth Edition

Heilbroner, Robert L.

THE WORLDLY PHILOSOPHERS: The Lives, Times and Ideas of the Great Economic Thinkers
Fourth Edition, Revised for the 1970's

"The first edition of this exciting book was published in 1953. In his introduction Heilbroner tells us that it is the 'search for order and meaning of social history that lies at the heart of economics. Hence [this] is the central theme of the book.' He achieves his objectives in a brilliant lucid manner. There is no more readable book available as an introduction to the great economists of the last 200 years. This fourth edition is substantially the same as earlier ones with the exception of Chapter 10 which deals with 'The modern world.' Because contemporary problems and trends change as well as what leading economists think about these problems, the content and emphasis of this chapter change with each new edition. Heilbroner has included 'A guide to further reading' with brief comments about the books that he suggests." —*Choice.*

☐ 21326 Biblio., Index $2.95

TOUCHSTONE

PAPERBACKS

SIMON AND SCHUSTER
630 FIFTH AVENUE
NEW YORK, N. Y. 10020

For complimentary examination copies, write on your departmental letterhead to School and Library Services, address above. Include details on course in which you are considering using title requested, and advise us where you have seen announcement.

JOB OPENINGS FOR ECONOMISTS

A bimonthly listing of job vacancies

Available only to AEA members and to organizations and departments of economics who agree to list their openings.

I would like to subscribe to *Job Openings for Economists* and enclose a check for one year.

1975 Subscription Rates—U.S.A., Canada, and Mexico (first class): $6.00 junior member, $12.00 others; foreign airmail: $12.00 junior member, $18.00 others.

Name ______________________________

Address ______________________________

____

Check one:

☐ I am a member of the American Economic Association.
☐ I am not a member but would like to become one. My application is enclosed.
☐ (For organizations) We agree to list our vacancies in JOE.

THE AMERICAN ECONOMIC ASSOCIATION
1313 21st Avenue South
Nashville, Tennessee 37212

New Books in Economics

Competition Policy
European and International Trends and Practices

A. E. Walsh and John Paxton

An important up-to-date reference for all those interested in the study of marketing goods, both with the European Economic Community and the remainder of the world's competitive economies.

Published 1975 200 pages $19.95, cloth

Development, Underdevelopment and the Law of Value
A Marxist Analysis

Geoffrey Kay

Considers the problem of underdevelopment through an analysis of the law of value as proposed and elaborated by Karl Marx.

Published 1975 208 pages $13.95, cloth

Economic Policies and Social Goals
Aspects of Public Choice

A. J. Culyer, editor

Presents a closely integrated series of original essays on substantive problems in social policy.

Published 1975 308 pages $25.00, cloth

Economic Theory and the Developing Countries
A. K. Dasgupta

Emphasizes the historical context of economic development and looks at the problems of development as they appear to the developing nations themselves.

Published 1975 250 pages $14.95, cloth

Inflation, Saving, and Growth in Developing Countries
A. P. Thirlwall

Brings together for the first time the role of monetary factors in the development process and presents old and new empirical evidence of the dependence of growth on saving and investment.

Published 1975 272 pages $19.95, cloth

Negotiating on Non-Tariff Distortions of Trade
The EFTA Precedents

Robert Middleton

Presents the methods developed by the EFTA for elimination of non-tariff distortions of trade and provides valuable guidance for the conduct of trade negotiations at multilateral and regional levels.

Published 1975 208 pages $18.95, cloth

The New Mercantilism
Harry G. Johnson, editor

Essays by leading economists examine the history of mercantilism, the state of mercantilism today, foreign trade and investment, regional economics, and the current monetary crisis.

Published 1975 128 pages $17.95, cloth

World Monetary Crisis
Peter Coffey

An analytical and critical examination of the world monetary order stressing both political and economic factors.

Published 1975 128 pages $14.95, cloth

St. Martin's Press
P.O. Box 5352
New York, New York 10017

Join the American Economic Association

Membership includes:

- a subscription to both *The American Economic Review* (quarterly) plus May, *Papers and Proceedings* and the *Journal of Economic Literature* (quarterly)
- annual ballots and nomination forms for elected officers of the Association
- preregistration material and announcement of the annual meeting of the Allied Social Science Associations
- announcements of special interest to members, such as the life insurance program started in 1973; new book offerings, etc.

Application for Membership

Please enroll me as a member of the American Economic Association. 1975 annual rates: Regular—$23; Family, two or more living at same address, second membership without publications—$2.30; Junior, *registered student $11.50/payable yearly; for a maximum of three years; Subscribing—$27.50; Contributing—$57.50; Life—$460. All dues payable in U.S. currency. Memberships are not entered until payment has been received. (Under the terms of a resolution passed at the annual meeting of members of the American Economic Association on Dec. 29, 1973. $18.40 of the annual membership dues is considered as payment for the AER and JEL; the remainder is considered a contribution to exempt functions of the Association.)

Name in full ________________________________

Address ________________________________

____
City State/Country Zip/Post Code

Endorsed by (AEA member) ________________________________

Below for Junior Members Only

* I certify that the person named above is enrolled as a student at ________

____

NOTE: Payment and certification must be submitted yearly.

____
Authorized Signature

Please send with remittance to: American Economic Association
1313 21st Avenue South
Nashville, Tennessee 37212

INTEGRATING INCOME MAINTENANCE PROGRAMS

edited by IRENE LURIE

A Volume in the INSTITUTE FOR RESEARCH ON POVERTY MONOGRAPH Series

The result of a conference of academic and government experts held under the aegis of the Institute for Research on Poverty, the eleven papers in this volume can serve as a basic reference and guide to the nation's income support system, and will be fundamental in any effort to reform it. They analyze its deficiencies and examine ways of integrating its programs into a system that makes sense.

CONTENTS: *I. Lurie,* Integrating Income Maintenance Programs: Problems and Solutions. *L. J. Hausman,* Cumulative Tax Rates in Alternative Income Maintenance Systems. *B. A. Okner,* The Role of Demogrants as an Income Maintenance Alternative. *R. H. Haveman,* Earnings Supplementation as an Income Maintenance Strategy: Issues of Program Structure and Integration. *T. W. Mirer,* Alternative Approaches to Integrating Income Transfer Programs. *H. J. Aaron,* Alternative Ways to Increase Work Effort under Income Maintenance Systems. *M. Schmundt, E. Smolensky, and L. Stiefel,* When Do Recipients Value Transfers at Their Costs to Taxpayers? *M. K. Taussig,* The Social Security Retirement Program and Welfare Reform. *R. Munts,* Programming Income Maintenance: The Place of Unemployment Insurance. *T. R. Marmor,* Public Medical Programs and Cash Assistance: The Problems of Program Integration. *M. Krashinsky,* Day Care and Welfare. *I. Lurie,* Rationales for Interstate Differentials in Assistance Benefits and for State Supplementation.

1975, 400 pp., $17.50/£8.40

SIMULATING THE URBAN ECONOMY

by P. SMITH and W. I. MORRISON

No. 7 of MONOGRAPHS IN SPATIAL AND ENVIRONMENTAL SYSTEMS ANALYSIS

Although the use of input-output analysis is increasing rapidly, the obstacle presented by the cost in time and money of obtaining survey data acts as a hindrance to even greater application of the method. The possibilities offered by nonsurvey methods are consequently very important.

In this book the authors have applied the methods developed by a number of workers to data derived from the 1968 UK input-output tables, and they have then compared the results with empirically obtained data for the city of Peterborough for the same year. This critical comparison of previous methods for the production of nonsurvey tables is accompanied by a number of the authors' own modifications to the methods, which demonstrate that significant improvements can be made in the estimates. In particular an iterative approach is described which incorporates sectoral estimates of local intermediate inputs and outputs, and makes possible the widespread simulation of regional economic systems.

1975, 152 pp., $10.50/£4.00

Published by PION LTD. Distributed by Academic Press.

BOUNDARY CONDITIONS AND GLOBAL MANAGEMENT

by GUIDO LOERO

On the basis of an analysis of a wide variety of economic, political, sociological, and psychological data and theories the author sees the conditions of modern life as pushing society towards major new resolutions of its problems, based on a broader viewpoint than that of mere economic efficiency. With particular attention to the role and problems of management, he sketches the new patterns of management decision-making, new structures of goverment, and new cultural patterns of needs and values whose emergence will better enable society to meet the human needs of its members.

1975, 200 pp., $14.50/£7.25

N.B.: Postage plus 50¢ handling charge on all orders not accompanied by payment.
Price subject to change without notice.

ACADEMIC PRESS, INC.
A Subsidiary of Harcourt Brace Jovanovich, Publishers
111 Fifth Avenue, New York, N.Y. 10003
24-28 Oval Road, London NW1 7DX

New from Columbia

Latest Publications from the National Bureau of Economic Research

FOREIGN TRADE REGIMES AND ECONOMIC DEVELOPMENT

Jagdish N. Bhagwati and Anne O. Krueger, Editors

This ambitious research project sponsored by the National Bureau of Economic Research analyzes the phenomena of exchange control, liberalization, and economic development in less-developed countries. There will be ten volumes on individual countries plus a synthesis volume.

- **Volume I: Turkey.** Anne O. Krueger, Editor. cloth, $17.50; paper, $5.00
- **Volume II: Ghana.** J. Clark Leith, Editor $15.00
- **Volume III: Israel.** Michael Michaely $12.50
- **Volume IV: Egypt.** Bent Hansen and Karim Nashashibi $17.50
- **Volume V: Philippines.** Robert Baldwin $10.00

A Special Conference Series on Foreign Trade Regimes and Economic Development

HOUSING MARKETS AND RACIAL DISCRIMINATION
A Microeconomic Analysis

John F. Kain and John M. Quigley

Based on a sample of roughly 1200 St. Louis households, 34% of which were Negro households, this study permits extensive analyses on the behavior of blacks in the housing market, pointing up the distortions caused by racial discrimination.

Urban and Regional Studies, 3 $17.50

HUMAN CAPITAL
A Theoretical and Empirical Analysis, with Special Reference to Education
Second Edition

Gary S. Becker

In the three new essays in this updated volume a convenient formulation is provided for statistical estimation of the relation between the lag of earnings, rates of return to human capital, and the time spent in investing in human capital; a model of wealth maximization is developed that explains the decline of investments over time; and a new approach is presented to household behavior.

Human Behavior and Social Institutions, 5 $10.00

COLUMBIA UNIVERSITY PRESS

Address for orders: 136 South Broadway, Irvington, New York 10533

The American Economic Review

ARTICLES

J. R. CROTTY AND L. A. RAPPING
The 1975 Report of the President's Council of Economic Advisers: A Radical Critique

M. RUBINSTEIN Securities Market Efficiency in an Arrow-Debreu Economy

F. MODIGLIANI AND E. TARANTELLI
The Consumption Function in a Developing Economy and The Italian Experience

W. E. WEBER Interest Rates, Inflation, and Consumer Expenditures

R. DORNBUSCH Exchange Rates and Fiscal Policy in a Popular Model of International Trade

B. J. MOORE Equities, Capital Gains, and the Role of Finance in Accumulation

D. F. BRADFORD Constraints on Government Investment Opportunities and the Choice of Discount Rate

I. FRIEND AND M. E. BLUME
The Demand for Risky Assets

A. RAZIN AND U. BEN-ZION
An Intergenerational Model of Population Growth

J. M. BARRON Search in the Labor Market and the Duration of Unemployment: Some Empirical Evidence

R. DEACON AND P. SHAPIRO
Private Preference for Collective Goods Revealed Through Voting on Referenda

C. R. HULTEN Technical Change and the Reproducibility of Capital

G. R. FAULHABER Cross-Subsidization: Pricing in Public Enterprises

H. HORI Revealed Preference for Public Goods

SHORTER PAPERS: O. Hochman; G. Daly and J. F. Giertz; W. W. Brown and M. O. Reynolds; R. Carson; D. A. Worcester, Jr.; A. Bergson; D. A. Graham and J. M. Vernon; S. M. Besen and R. Soligo.

CALCUTTA * CENTRAL LIBRARY

DECEMBER 1975

THE AMERICAN ECONOMIC ASSOCIATION

Founded in 1885

- Published at George Banta Co., Inc., Menasha, Wisconsin.

- THE AMERICAN ECONOMIC REVIEW, including four quarterly numbers, the *Proceedings* of the annual meetings, the *Directory*, and Supplements, is published by the American Economic Association and is sent to all members five times a year, in March, May, June, September, and December.

Association membership dues for 1976, which include a subscription to both The *American Economic Review* and the *Journal of Economic Literature*, are as follows:

$25.00 for regular members with rank of assistant professor or lower, or with annual income of $12,000 or less;

$30.00 for regular members with rank of associate professor, or with annual income of $12,000 to $20,000;

$35.00 for regular members with rank of full professor, or with annual income above $20,000;

$12.50 for junior members (registered students). Certification must be submitted yearly.

Subscriptions (libraries, institutions, or firms) are $37.50 a year. Only subscriptions to both publications will be accepted. Single copies of either journal may be purchased from the Secretary's office, Nashville, Tennessee.

For foreign postage, add: PUAS countries, $2.10; all other countries and Canada, $2.50.

- Correspondence relating to the *Papers and Proceedings*, the *Directory*, advertising, permission to quote, business matters, subscriptions, membership and changes of address should be sent to the Secretary, Rendigs Fels, 1313 21st Avenue South, Nashville, Tennessee 37212. Change of address must reach the Secretary at least six (6) weeks prior to the month of publication. The Association's publications are mailed second class.

- Second-class postage paid at Nashville, Tennessee and at additional mailing offices. Printed in U.S.A.

Officers

President

ROBERT AARON GORDON
University of California, Berkeley

President-Elect

FRANCO MODIGLIANI
Massachusetts Institute of Technology

Vice-Presidents

MARTIN BRONFENBRENNER
Duke University
CHARLES L. SCHULTZE
University of Maryland
and The Brookings Institution

Secretary-Treasurer and Editor of the Proceedings

RENDIGS FELS
Vanderbilt University

Assistant Secretary-Treasurer

C. ELTON HINSHAW
Vanderbilt University

Managing Editor of The American Economic Review

GEORGE H. BORTS
Brown University

Managing Editor of The Journal of Economic Literature

MARK PERLMAN
University of Pittsburgh

Executive Committee

Elected Members of the Executive Committee

IRMA ADELMAN
University of Maryland
ROBERT HEILBRONER
New School for Social Research
ANDREW F. BRIMMER
Harvard University
PAUL M. SWEEZY
Monthly Review
CAROLYN SHAW BELL
Wellesley College
BURTON A. WEISBROD
University of Wisconsin, Madison

Ex Officio Members

KENNETH J. ARROW
Harvard University
WALTER W. HELLER
University of Minnesota

THE AMERICAN ECONOMIC REVIEW

GEORGE H. BORTS
Managing Editor

WILMA ST. JOHN
Assistant Editor

Board of Editors

BELA BALASSA
EUGENE F. FAMA
MARTIN S. FELDSTEIN
ANN F. FRIEDLAENDER
STEPHEN M. GOLDFELD
ROBERT J. GORDON
ROBERT E. HALL
BENT HANSEN
JAMES R. MELVIN
WILLIAM D. NORDHAUS
LEONARD A. RAPPING
STEPHEN RESNICK
ANNA J. SCHWARTZ
JEROME STEIN
JOSEPH E. STIGLITZ
S. C. TSIANG
FINIS WELCH
MARINA V. N. WHITMAN

• Manuscripts and editorial correspondence relating to the regular quarterly issue of this REVIEW should be addressed to George H. Borts, Managing Editor of THE AMERICAN ECONOMIC REVIEW, Brown University, Providence, R.I. 02912. Manuscripts should be submitted in duplicate and in acceptable form and should be no longer than 50 pages of double-spaced typescript. A submission fee must accompany each manuscript: $15 for members; $30 for nonmembers. *Style Instructions* for guidance in preparing manuscripts will be provided upon request to the editor.

• No responsibility for the views expressed by authors in this REVIEW is assumed by the editors or the publishers, The American Economic Association.

• Copyright American Economic Association 1975.

December 1975

VOLUME LXV, NUMBER 5

Articles

The 1975 Report of the President's Council of Economic Advisers: A Radical Critique *James R. Crotty and Leonard A. Rapping* 791

Securities Market Efficiency in an Arrow-Debreu Economy *Mark Rubinstein* 812

The Consumption Function in a Developing Economy and The Italian Experience *F. Modigliani and E. Tarantelli* 825

Interest Rates, Inflation, and Consumer Expenditures *Warren E. Weber* 843

Exchange Rates and Fiscal Policy in a Popular Model of International Trade *Rudiger Dornbusch* 859

Equities, Capital Gains, and the Role of Finance in Accumulation *B. J. Moore* 872

Constraints on Government Investment Opportunities and the Choice of Discount Rate *David F. Bradford* 887

The Demand for Risky Assets *Irwin Friend and Marshall E. Blume* 900

An Intergenerational Model of Population Growth *Assaf Razin and Uri Ben-Zion* 923

Search in the Labor Market and the Duration of Unemployment: Some Empirical Evidence *John M. Barron* 934

Private Preference for Collective Goods Revealed Through Voting on Referenda *Robert Deacon and Perry Shapiro* 943

Technical Change and the Reproducibility of Capital *Charles R. Hulten* 956

Cross-Subsidization: Pricing in Public Enterprises *Gerald R. Faulhaber* 966

Revealed Preference for Public Goods *Hajime Hori* 978

Shorter Papers

Market Equilibrium Versus Optimum in a Model with Congestion: Note *Oded Hochman* 992
Externalities, Extortion, and Efficiency *George Daly and J. Fred Giertz* 997
A Model of IQ, Occupation, and Earnings *William W. Brown and Morgan O. Reynolds* 1002
On Monopoly Welfare Losses:
Comment *R. Carson* 1008
Comment *Dean A. Worcester, Jr.* 1015
Reply *Abram Bergson* 1024
The Economics of the Network-Affiliate Relationship:
Comment *Daniel A. Graham and John M. Vernon* 1032
Reply *Stanley M. Besen and Ronald Soligo* 1037
Notes 1039
Titles of Doctoral Dissertations 1047

The following Statement of Ownership, Management, and Circulation is provided in accordance with the requirements, as contained in 39 U.S. Code 3685. The *American Economic Review* is owned, managed, and published by the American Economic Association, a nonprofit educational organization, located at 1313 21st Avenue So., Nashville, Davidson County, Tennessee 37212. The Managing Editor is Professor George H. Borts, *American Economic Review*, Brown University, Robinson Hall, Providence, R.I. 02912. During the preceding 12 months the average number of copies printed for each issue was 27,400, the average paid circulation, 24,752; the average free distribution, 327; the average number of copies distributed, 25,079. Corresponding figures for the last issue before filing: 27,500 total number copies printed; 25,642 total paid circulation; 327 free copies distributed; 25,969 total distributions.

The 1975 Report of the President's Council of Economic Advisors: A Radical Critique

By JAMES R. CROTTY AND LEONARD A. RAPPING*

Capitalism by its very nature is an exploitative system. During periods of recession and depression, however, the burdens of capitalism on working people are intensified and brought into stark relief. Economic and political inequality grows, poverty becomes more pervasive, working conditions deteriorate, work relations become more dictatorial, economic insecurity increases, and the condition of blacks and other minorities as well as that of women worsens. All of these injustices accompany the semidepression of 1974–75, which is the most severe economic crisis since the 1930's.

The Employment Act of 1946 is generally assumed to have affirmed the government's commitment to relatively low unemployment and to the maintenance of reasonable price stability.[1] Yet 1974 was a year in which the unemployment rate was high and grew rapidly toward year's end. By May of 1975, unemployment reached its highest level in almost thirty-five years. During the year 1974, real wages fell; real consumption was stagnant; and the United States was experiencing one of the worst peacetime inflations in its history. In light of these events, the 1975 Economic Report of the President is an exceptionally important document because it represents an official explanation of why the government has either refused or been unable to fulfill its alleged promise of 1946.

The Report contends that the economic problems of 1974 resulted from a decade-long series of mistaken macropolicies which, in conjunction with oil and food price complications, left the policy makers in 1973–74 with no reasonable choice other than to precipitate a serious recession. In other words, with better policy judgment and a little more luck, the necessity of severe economic retrenchment could have and should have been avoided. We, on the other hand, will argue that the current economic crisis is systemic in nature and that the macropolicies of the past decade were rational and perhaps unavoidable given the political-economic objectives and constraints facing the macropolicy authorities. In this review, therefore, we will offer an explanation of the key events of 1974 that differs from the one presented in the Report.

---

* Economics department, University of Massachusetts. This review was written in mid-summer of 1975. We have benefited from the helpful comments of Sam Bowles, Richard Edwards, Tom Riddell, and Helen Sootin Smith. We are especially indebted to Raford Boddy who helped develop many of the ideas presented here. We acknowledge the intellectual support of our colleagues in the Union for Radical Political Economics, but we caution the reader not to assume that there is general acceptance within URPE of the views expressed in this paper.

[1] The exact wording of the 1946 Act as well as the congressional debate before its passage makes clear that the government did not commit itself to *full* employment. The Act itself was ambiguous, calling for "maximum employment and purchasing power," subject to the condition that these vague objectives not jeopardize the maintenance of the existing system of "free competitive enterprise." And it was chosen instead of an alternative bill under which the government really was to guarantee "full employment" for persons able and willing to work.

I. The Report: Description Without Explanation

The Report argues that the recession of 1974 and the accompanying rise in unemployment were caused in the first instance by the restrictive macropolicies of 1973 and 1974. These policies were required, in the view of the Report, in order to fight the inflation and correct the economic distortions which existed in this period. The unusual importance attached to the fight against inflation as well as the severity of the restrictive policies undertaken to combat it are basically explained by the inflationary expectations and tendencies which had become embedded in the economy after more than eight years of consistent inflationary pressure. In the logic of the Report, therefore, whatever explains the pattern of inflation since 1965 is also the key to understanding the recession of 1974 and beyond.

The explanation of inflation offered in the chapter on inflation and throughout the Report focuses primarily on the role of macropolicy. The opening paragraph in this chapter instructs us as follows:

> Without neglecting specific features, the U.S. inflation since the mid-1960's can nevertheless be analyzed in terms of a general conception of the inflationary process that emphasizes the role of monetary and fiscal policies, *and the decision-making process that leads to these policies.* [p. 128] (emphasis added)

The explanation begins with the obvious proposition that inflation develops when aggregate demand exceeds aggregate supply and the less self-evident proposition that inflation is a cumulative process—the longer demand exceeds supply, the greater the inflationary pressure. The Report argues: "Any explanation of inflation must therefore come to grips with the question of why, about 1965, aggregate demand started to grow so much faster in nominal terms than real output, and why it has continued to grow at a faster rate subsequently" (p. 132). It argues further that the growth of demand was caused primarily by overexpansionary macropolicy. Since fiscal policy is assumed incapable of sustaining demand growth in the absence of accomodating monetary policy, the explanation reduces to an observation of the *fact* that "the Federal Reserve System did not prevent this sustained period of steepening inflation" or that "the regulation of the money supply has not prevented these undesireable price trends" (pp. 133 and 134).

The Report never really tells us why the Federal Reserve permitted this decade of inflation. The monetary authorities were hardly unaware of the problem. In fact, the Report notes that the Federal Reserve did tighten policy on three occasions during this period—1966–67, 1969–70, and 1973–74. It asserts, however, that certain "pressures" were placed on the monetary authorities to return to expansionary policies. These pressures are said to have become increasingly strong in late 1974, but their source is not identified precisely. One motive suggested is " . . . the effort of [monetary] policy makers to play safe against recessions or at least to postpone them and to promote a very rapid rate of cyclical expansion in the advanced stages of the recovery after the recession of 1970" (pp. 135–36). Are we to imagine William McChesney Martin and Arthur Burns fighting at all costs to prevent recession, whatever the inflationary or balance of payments consequences? Surely the pressures on the Federal Reserve System in 1966–67 and 1970 were something other than or besides a simple abhorrence of unemployment.

In spite of the fact that the Report stresses that any explanation of the key events of 1974 and the inflation of the preceding decade must be based on an analysis of macropolicy and "the decision-making process that leads to these policies," it does not follow its own dictum. It presents no explicit theory of the macropolicy decision-making process. The Report never even clearly specifies a set of macropolicy objectives and constraints, let alone a political theory to be used as the analytical foundation for understanding the process of macropolicy decision making. Scattered throughout are a mélange of such policy objectives as price stability and income and employment maximization which are implicitly assumed to be in everyone's best interest. These objectives are normally associated with a relatively simple if not simplistic pluralist political theory. But such a theory is clearly inadequate for understanding the political realities of a capitalist system characterized by the existence of classes whose conflicts influence or dominate the political process. While it may pass muster in high school civics texts, naive pluralism won't do as a guide to a world in which giant corporations dominate the domestic and international economy, and where their owners and managers generally exercise decisive influence in political institutions and their supporting ideology envelops virtually all public political debate.[2] Lacking a realistic underlying theory of the state, therefore, the Report's explanation of the economic crisis finally turns out to be mechanical, superficial, and behaviorally empty. It must be judged inadequate under its own ground rules.

For the purpose of this review, we accept the Report's *description* of the events of 1974 and will focus on the question of *why* 1974 turned out to be an economic disaster. The reader should not expect a detailed critique of the Report nor a consideration of the numerous controversial technical issues which it raises. Rather, our discussion will center on broad questions concerning the underlying structure of the American economy, and the economic and political factors which explain the formation of monetary-fiscal policy. Consistent with the Report, we assume that the events of 1974 are only fully comprehensible when viewed as the culmination of an organic historical process.

We begin by sketching the essential components of a theory of macropolicy consistent with our understanding of Marxian social theory and a derivative theory of the state. This theory of policy is based on class and nation-state conflict and stresses contradictions in both the domestic and international economic systems. We then use this theory to interpret or explain important economic developments in the 1965–73 period that set the stage for the inflation and recession-depression of 1974–75. Finally, in the context of the current crisis, we speculate on the political-economic prospects for the rest of the decade.

II. The Foundation of a Theory of the Crisis

The history of American economic development since World War II and of the macropolicy of this era is best understood in terms of several fundamental conflicts. This analysis will stress two of these. Domestically, there exists a basic conflict between capital and labor which leads to a struggle over the division of the national product. Internationally, a fundamental conflict exists among the various capitalist nation-states over export markets, access to raw materials, control of foreign investment and—in a very general sense—each country's place in the hierarchy of international economic, political, and military relations. This competitive struggle takes place within a broader conflict between the third world underdeveloped nations and the developed capitalist nations. The presence of developed or developing countries outside the capitalist orbit constrains intercapitalist rivalry and offers alternatives to underdeveloped nations.

---

[2] We do not at all mean to suggest that the absence of an explicit, well-developed theory of policy or theory of political economy in the Report implies an equivalent absence of political sophistication or lack of ideological commitment in the Council members who wrote it. On the contrary, this year's Council is one of the most conservative in history. Alan Greenspan is a disciple of Ayn Rand and perhaps the most openly ideological Council chairman ever. The political sophistication of this or any Council should be judged not by the apparent quality of the political economy of the Report, but by the degree to which the Report accomplishes its real purpose: to justify, not critically analyze, the policies of the Administration; to explain away, rather than explain, the economic problems of the period; and to praise, not question, capitalism as an economic system. The political and ideological bent of the Council will be determined by the basic policy thrust of a particular administration. For example, a conservative Council headed by Greenspan seems ideally suited for purging the economy of inflationary excesses by means of a severe recession-depression.

A. *Class Conflict and Macropolicy*

While it is undoubtedly true that the working class has achieved important political victories in modern American history, the process of macropolicy formulation in the post-World War II era has been dominated by the interests of the capitalist class.[3] Organized labor has represented a constraint limiting the policy choices available to capital, but by no means has it been an equal partner in the formation of macropolicy. The main objective of policy has been to protect and extend corporate profits. John Gurley expressed this point well in this *Review:*

> The ultimate and overriding objective of all capitalist governments is the preservation and strengthening of capitalism—that is, of the private enterprise system. In our own economy, this means that successive Administrations, above all else, have looked out for the welfare of . . . the few hundred industrial, commercial, and financial giants that define and shape the U.S. economy. Since the welfare of these huge corporations is best measured by their profits, the basic concern of U.S. Administrations has been the protection and promotion of these corporate profits . . . [F]ull employment, welfare programs, and price stability, are, in the final analysis, judged by the extent to which they are compatible with private profit making. [p. 20]

Capital's control of macropolicy and its quest for maximum corporate profits would be a secondary issue were the impact of policy always class-neutral. The 1975 Report implicitly assumes, as did its predecessors, that policy choices exist that are optimal for both capital and labor. In particular, this harmonious view assumes that labor's key objective, sustained full employment, and capital's objective, maximum profits, are compatible.[4] According to this reasoning, reaching the twin objectives of full employment and maximum profits involves mainly technical problems, and it is assumed that the basic structure of the economy is consistent with the simultaneous achievement of both goals. Neoclassical and Keynesian theorists lent considerable support to this view because they implicitly and erroneously assumed that the unique conjunction of historical events after World War II, which facilitated worldwide capitalist expansion and growth under American tutelage, would last forever. And the Employment Act of

---

[3] Supporting the important assertion that capital controls macropolicy formation would take us beyond the scope of this paper. We have little doubt, however, that policy makers are both sociologically and ideologically more closely tied to capital than labor. Moreover, the dominance of capital in the policy formation process is clear to us from our reading of the historical "case studies" reported by those who have examined macropolicy formation. The best and most informative of these studies has been done by Herbert Stein. For an analysis and some empirical tests of the proposition that capital dominates macropolicy formation, see Crotty.

[4] Within the constraints of *existing* institutions or relations of production some policy objectives, such as the avoidance of economic collapse, are in the interests of both classes. The reader who questions the primacy of the job issue in the objectives of organized labor should consult the trade union literature and AFL-CIO policy positions during every recessionary period.

1946 seemed to commit the state to balanced full employment growth while Keynesian theory indicated that it could be achieved.

In the final chapter of the *General Theory*, Keynes declared that capitalism had but one major problem which could be permanently resolved through sensible government policy. That problem of course was that aggregate demand tended to fall short of aggregate supply at full employment leading to recessions or depressions which were not self-correcting. He wrote:

> If our central controls succeed in establishing an aggregate volume of output corresponding to full employment . . . , then there is no objection to be raised against the classical analysis of the manner in which private self-interest will determine what in particular is produced, in what proportions the factors of production will be combined to produce it, and how the value of the final product will be distributed between them. [pp. 378–79]

Keynes' key assumption is that the market system coordinating firm with firm, sector with sector, and factor with factor, works reasonably well and quickly when the government sets aggregate demand at the level consistent with full employment. In other words, the market system can handle all the tasks of micro-balancing. Moreover, once government provides full employment through macropolicy, the international sector is no longer an overriding policy constraint: "If nations can learn to provide themselves with full employment by their domestic policy . . . there would no longer be a pressing motive why one country need force its wares on another or repulse the offerings of its neighbor . . . with the express object of upsetting the equilibrium of payments so as to develop a balance of trade in its own favor" (p. 382).

An alternative to the Keynesian view of macropolicy, one based on capital-labor conflict, maintains that the basic structure of the economy precludes the simultaneous attainment of labor's and capital's objectives. The existence of a conflict between full employment and profit maximization is supported by empirical evidence from the past twenty-five years. The typical pattern of the post-World War II business cycle shows that the profit share of income and even the absolute level of real profits have fallen in the latter phase of the expansion. This full employment profit squeeze helps explain both the failure of capitalist-oriented macropolicy to sustain full employment prior to 1965 and the consequences of full employment during the Vietnam War period of the late 1960's.

The primary problems with sustained full employment from the perspective of the capitalist class are that it creates conditions which lead to the continuous decline of profits and profit margins, to an increase in labor's share of national product, to domestic financial instability, and to the erosion of the international financial institutions through which the American empire was developed and is maintained. Intermittent recessions are a necessary condition for the maintenance of satisfactory profits and the health of financial institutions. The tradeoff between profits and full employment enables the state to partially mediate this capital-labor conflict through macropolicy. Control of the policy-making apparatus therefore becomes central in determining the influence of macropolicy on the outcome of this conflict. The state cannot be expected to pursue full employment since those who formulate macropolicy are primarily concerned with the welfare of capital and the preservation of capitalist institutions.

B. *Profits, Wages, Productivity, and the Cycle*

In the absence of wage-price controls,

full employment and high profits are incompatible. This contradiction is a cornerstone of Marxian business cycle theory.[5] Consider the economy emerging from a recession and entering the early phase of an expansion. Wages are rising moderately, but productivity growth is very rapid, so that stable or moderately rising prices yield increasing profits and profit margins. Capacity utilization is rising but is not yet so high that shortages develop. Credit is cheap and readily available, but credit demand is moderate because cash flow is sufficient to accommodate most business needs. Investment is rising, business and consumer confidence is strong, and the stock market is buoyant. Trade and payments accounts are improving. Employment is growing and with a lag the rate of unemployment begins to decline from its recession highs.

When relatively full employment is approached, however, a series of distortions develop which, in Wesley Mitchell's words, "cause a slow accumulation of stresses within the balanced system of business . . . that ultimately undermine the conditions on which prosperity rests" (quoted in Geoffrey H. Moore, p. 7). For Marx, the most important of these stresses derives from the continuous conflict of capital and labor over wages and the conditions of work. Marx argued that the division of national product depended in part on class struggle. Capital-labor conflict over wage and profit shares, coupled with the relentless drive for capital accumulation, produced, according to Marx, cyclical booms and busts.

The reduction of the reserve army of the unemployed which accompanies the expansion, strengthens the bargaining position of the working class in its labor market confrontation with capital. Workers can then struggle successfully for higher wages. In the latter part of the expansion, moreover, the rate of increase in productivity decreases and consequently unit labor costs rise rapidly. While the growth in unit labor costs drives corporations toward labor saving capital equipment, the resulting productivity gains are only realized in the early phase of the next expansion when the new capital is actually installed. Although developing "technical" bottlenecks and inefficiencies partially explain the productivity slowdown, knowledgeable observers of the labor scene have pointed directly to an increasingly obstreperous labor force as an explanatory factor. Full employment affects not only wage rates, but also quit rates, strikes, and labor militancy in general.[6] It has been a longstanding observation that such favorable conditions for labor evoke less effort on the shop floor.[7] For example, a plant superintendent testifying before the Commission on Labor in 1904 had the following to say:

> Five years ago men did not restrict their output, union or non-union men, because they wanted to hold their jobs and a man would do anything and all he was told to do. Now a man knows that all he has to do is to walk across the street and get another job at the same pay. During the hard times we took contracts at any price we could get, and in some places and cases men were driven at high speed to get this work out, so as to lose as little money on it as possible. Men will not keep up that speed rate in these days. [as quoted in Mitchell, p. 113]

In this phase of the expansion, inflation accelerates, but not enough to prevent profit margins from deteriorating. Although there is no unique and universally accepted hypothesis that explains why prices do not keep up with costs, the fact that they do not is of extreme importance. For profits and profit margins end up being squeezed toward the top of the boom. The termination of the expansion and the rebuilding of the reserve army is required to relax this profit squeeze. As Marx put it:

---

[5] A fuller discussion of the propositions contained in this section can be found in Raford Boddy and Crotty (1975a).

[6] For evidence on the relationship between labor market conditions and strike activity, see Douglas Hibbs.

[7] In Marxian economics a useful distinction is made between labor and labor power. The distinction is explored in a recent review of human capital theory by Samuel Bowles and Herbert Gintis.

> If the quantity of unpaid labor supplied by the working class, and accumulated by the capitalist class, increases so rapidly that its conversion into capital requires an extraordinary addition of paid labor, then wages rise, and, all other circumstances remaining equal, the unpaid labor diminishes in proportion. But as soon as the diminution touches the point at which the surplus labor that nourishes capital is no longer supplied in normal quantity, a reaction sets in: a smaller part of revenue is capitalized, accumulation lags, and the movement of rise in wages receives a check. The rise of wages therefore is confined within limits that not only leave intact the foundations of the capitalist system, but also secure its reproduction on a progressive scale. [p. 620]

Empirical evidence from the post-World War II era supports the hypothesis that there is indeed a full employment profit squeeze.[8] Profits tend to peak, both in relative and absolute terms, approximately halfway through the typical expansion. Profitability is eroded during the latter part of this phase as labor markets tighten and continues to decline during the contraction. This pattern is discernable in each of the post-World War II cycles. However, the expansion of the 1960's is the only one in which there was prolonged full employment. It therefore provides us with the best opportunity to test the full employment profit squeeze hypothesis. The evidence of this period dramatically supports the hypothesis, whether profitability is measured as a share, as a rate of return, or in dollars. Net profits before taxes as a percent of gross product originating in nonfinancial corporations rose steadily from 14.5 in 1961 to a peak of 17.0 in 1965, then fell steadily to a postwar low of 9.8 by 1970. Compensation of employees, on the other hand, declined from 65.1 percent in 1961 to 62.6 in 1965, rising thereafter to 65.1 in 1969 and 67.2 in 1970. As reported by Nordhaus, the rate of return on nonfinancial corporate capital rose from 11.8 percent (before taxes) in 1961 to 16.3 percent in 1965; it then fell steadily to a post-World War II low of 9.1 percent in 1970. Before tax corporate profits for all industries in current dollars rose from $50.3 billion to $82.4 billion between 1961 and 1966. They were $79.8 billion and $69.2 billion in 1969 and 1970, respectively. In real terms, they declined steadily and strongly from 1966 through 1970.

The full employment squeeze on profits means that continuous expansion would undermine the accumulation of profits. It is the profit squeeze, not inflation per se, which motivates capitalists to eventually demand restrictive macropolicies during periods of full employment.[9] The fact that profits are lowest in recession has often been used to argue the preposterousness of the charge that recessions are desired by capital. Indeed they would prefer to avoid them. But the recession is a necessary condition for establishing the basis of the highly profitable first phase of the expansion and for avoiding the highly unprofitable consequences of sustained full employment. Recessions also provide the political context within which the government grants fiscal incentives to capital. Under the impact of recession-induced corporate tax reductions, the post-World War II period has witnessed a secular decline in effective corporate tax rates.

---

[8] Evidence in support of the full employment profit squeeze hypothesis may be found in Boddy and Crotty (1975a), Albert Burger, Thor Hultgren, and William Nordhaus.

[9] Michał Kalecki in his seminal 1943 paper argued that periodic recessions were essential to the maintenance of capitalist institutions. He stated that "... under a regime of permanent full employment, the 'sack' would cease to play its role as a disciplinary measure." For Kalecki it was not the conflict between high profits and full employment that necessitated recessions, but rather the fact that full employment would undermine "discipline in the factory" and that the "class-consciousness of the working class would grow."

Of course other problems also develop during the expansion. The deterioration of profitability in the latter phase of the expansion means that corporations turn increasingly to external sources to finance investment in plant, equipment, and inventories. Rising prices and the fear of shortages accelerate inventory accumulation, while inventory profits and depreciation at historical costs increase the effective tax rate on corporate profits. Debt accumulates just when interest rates are highest. Moreover, serious balance of trade problems develop as the rising price of *U.S.* products retards exports and, aided by strong aggregate *U.S.* demand, stimulates imports.

Eventually of course lower profits and associated stresses restrain production and investment and thus, before long, the expansion tends to abort. Macropolicy can work by either acquiescing in this tendency toward contraction or can extend the expansion for months or even years beyond its "natural" life. In the 1945–65 period, the government reacted to the decline in profits, the rise in inflation, and the problems in the international sector as we would expect—by reinforcing the recessionary pressures developing in the private sector.

The economic function of recession in a capitalist economy is to create the preconditions for a new expansion of profits. By taking away the jobs of millions of people, and by threatening the jobs of millions of others, recessions reduce worker demands. Unit labor costs are brought under control, the decline in profit margins is eventually stopped, and inflationary pressure weakens. During recessions inventories are cut, loans are repaid, corporate liquidity positions are improved, and the deterioration in the balance of payments is reversed. All the statements of Keynesian economists to the contrary notwithstanding, recessions are inevitable in the unplanned economy of the United States because they perform an essential function for which no adequate substitute has thus far been available.

C. *World Capitalism, U.S. Hegemony, and Macropolicy*

The preceding analysis assumes that when capital controls the process of macropolicy making, the primary concern of policy is the profitability of major corporations. However, corporate profits—domestic as well as foreign—are fundamentally dependent upon the condition of the international economy and the relative position of the United States therein. The state, therefore, must formulate macropolicies which support and sustain world capitalism as a system, and, simultaneously, attempt to maximize the strength of the United States vis-à-vis its capitalist rivals. Pursuit of these two international objectives sometimes leads to serious policy contradictions.

Relations among independent, capitalist nation-states generally reflect both the harmony of interest binding together the fortunes of these states and the much more important intercapitalist rivalries inherent in these relations. This harmony-conflict dialectic manifests itself in every historical period. The periods of international harmony are usually those in which one dominant nation-state orders intercapitalist relations and suppresses conflict. International stability as we have known it seems to require a single hegemonic state.[10]

---

[10] This view is common in the Marxist literature, but for the reader who finds this literature unbecoming, the

It is the function of the dominant state to provide reasonably unfettered access to its markets, to provide credit, to maintain free trade, and to assure a liberal flow of capital and aid. For a world in which powerful noncapitalist states exist and in which third world revolutionary forces are increasingly insistent on a change in the social relations of production, the dominant capitalist state must also provide the military and political might and leadership to contain socialism and to prevent revolution. Without a dominant power the exchange relations of the capitalist world are basically unstable and break out in destructive, competitive rivalries. Trade wars, rising tariffs, import quotas, exchange controls, and barter arrangements are among the more important manifestations of disintegrating international exchange relations. The collapse of trade and credit relations in the 1930's and the uncertainties of the 1970's are two important historical examples of the instability of the world economy in the absence of a dominant state. Indeed, the Report expresses surprise and relief that protectionism was not more virulent in 1974.

The basis for stability in the postwar international system prior to August of 1971 was American political, economic, and military dominance. Within the framework of the Bretton Woods agreement, exchange rates were stabilized and the dollar served as both an international money and a reserve currency. Markets for both goods and direct investments as well as sources of cheap raw materials were protected and expanded by American military and economic assistance, which was provided directly as well as indirectly through American-dominated international agencies. Third world revolutionary forces were suppressed or contained by the unrelenting activities of *U.S.* agencies such as the Pentagon, the CIA, and the AID.

It is crucial to understand that the relative stability and prosperity of the *U.S.* economy in the period from 1946 to 1969 was based on the stability and order of the international capitalist system and the paramount position of the United States in that system. Foreign profits of *U.S.* corporations rose steadily throughout the period, both absolutely and as a percent of total corporate profits. The United States was guaranteed a supply of cheap raw materials and it gained control of an increasing proportion of world oil supplies. The entire nonsocialist world was kept "open" for *U.S.* trade and foreign investment. The policies of third world countries in particular were shaped so as to maximize the profitability and safety of *U.S.* interests. Exchange rates were not only kept stable, but the *U.S.* dollar was increasingly overvalued as time went on, providing cheap imports and facilitating the accumulation of foreign capital assets. The reserve position of the dollar not only meant that the economy was insulated from retrenchment shocks required by short-run balance-of-payments problems, but that the United States could finance much of its overseas military, economic, and political activities by printing dollars.[11] In general, the ability of the United States to control its international environment to a greater degree than in any other period in its history created international stability in the context of American dominance which was the sine qua non of the relative domestic sta-

---

recent book by Charles P. Kindleberger on the Great Depression will be enlightening. Kindleberger prefers to use the term leadership rather than hegemony, but regardless of terminology, the point is that stability in the international capitalist system requires an enforcer.

[11] Between 1950 and 1973, the United States directly exported (net) $96 billion of long-term capital and expended $73 billion on military and economic aid. Of this total, $78 billion was financed by earnings on the export of goods and services, while $91 billion was financed by short-term capital and gold flows. (Source: Landsberg, Table 1.) For a fuller discussion of this process, see Paul Sweezy and Harry Magdoff.

bility and the prosperity of this era. The economic chaos of 1974 is in large part the result of the end of the total dominance of the United States and the international disorder which accompanied it.

Post-World War II international stability facilitated world capitalist integration and unprecedented economic growth, as witnessed by increasing world trade, extensive direct investments, and the rise of the multinational corporation. By pre-1939 standards the separate economies of the Western capitalist countries and Japan also experienced substantial growth and stability. But this system was destined to self-destruct. The growing strength of the European Economic Community and Japan greatly increased world capitalist competition, undermining the basis of international order—*U.S.* hegemony. And the increasing strength of third world revolutionary forces eventually led to the American defeat in Southeast Asia, a major historical event that further eroded *U.S.* control. Finally, a cornerstone of stability collapsed in August of 1971 when the Bretton Woods system came to an end.[12] The power vacuum created by these developments has yet to be filled.

Pressures on macropolicy emanating from the international sector can be placed in two general categories: 1) the "normal" month-to-month and year-to-year dynamics of trade, investment, and financial capital flows; and 2) the inevitable wars and lesser military skirmishes which arise as a natural result of the United States' role as capitalist policemen of the third world and defender of the international capitalist system against the encroachment of socialism. Between 1945 and 1965 there existed a rough correspondence between domestic macropolicy requirements and the requirements of the international system. The behavior of the economy over the cycle generates simultaneous pressures to relax the economy for both domestic and "normal" international objectives. The refusal of the macropolicy authorities to follow expansionary policies during the recession of 1957–59 is an important historical example of this coincidence of objectives. There is, of course, no systematic reason to expect the changes in military expenditures necessitated by *U.S.* international objectives to coincide with domestic economic needs. Yet, the military expenditures on World War II and on the Korean War did help to revive a sagging economy.

The first dramatic conflict among these policy objectives occurred in 1965. Late that year, as 4 percent unemployment was attained, profit margins peaked and the trade surplus began to deteriorate. Both domestic and normal international interests required a recession and the Council of Economic Advisers called for more restrictive policies.[13] The Johnson Administration, however, with the approval and support of corporation owners and managers, had already made the fateful judgment that a military victory over the revolutionary forces in Vietnam was essential to the maintenance of the world capitalist system and the protection of the American place in it.

III. Working Out the Contradictions: 1965–74

Having sketched the foundations of a theory of the crisis, we now use these ideas to interpret and explain the important economic developments beginning in the critical year 1965 that set the stage for the inflation and recession-depression of 1974–75.

---

[12] For a more complete exposition and analysis of the events which ended *U.S.* dominance, see the interesting paper by Arthur MacEwan.

[13] Arthur Okun, then member of the Council of Economic Advisors, states that: "the President was told by his economic advisers at the end of 1965 that a general increase in income taxes was desirable to avoid excess demand" (p. 70).

A. *The Period 1965–71*

The administration's decision to escalate the war was a momentous gamble. Unless the government compensated for the sharp rise in military expenditures with either increased taxes or greatly reduced domestic spending, the war-related expenditures would eventually generate domestic economic instability, would place a severe strain on international financial institutions, and would jeopardize the special status of the United States in the world capitalist system. But the administration was constrained. It was unable to reduce domestic spending because of explosive unrest in the black urban areas, and it was unwilling to raise taxes because of fear that this would fuel incipient political opposition to war policies.[14] Except for the delayed and ineffective tax surcharge of 1968, the Johnson Administration chose to let inflationary war finance run its course. The Federal Reserve, on its part, started to squeeze money markets in late 1966 in an attempt to reduce demand pressures. But they almost initiated a financial collapse in the process and relented under pressure from the corporate sector.

The gamble failed. Yet the fact that it was undertaken, given the enormous political-economic costs involved, indicates how important the defense of the empire was to the policy makers. No other single event speaks so eloquently and convincingly in support of the hypothesis that the international interests of *U.S.* corporations are of paramount importance in the objective function of macropolicy authorities.

By the end of 1968 the costs of the Vietnam War had become obvious. The domestic economy had been severely strained by long years of expansion and dangerous pressures were building on the world payments system. Profit margins were low, labor's share was growing, interest rates were high, inflation was accelerating, and the *U.S.* balance-of-payments position was deteriorating. Certainly it must have appeared that a recession lasting as long as three or four years would be required to curb the pressures and end the dislocations generated by eight years of continuous expansion and five years of continuous tight labor markets. In 1969, the incoming administration began to deal with these problems by means of classic recessionary policy.

The actual recession of 1969–70 was rather short, lasting only five or six quarters. Restrictive macropolicy was eased in 1970 because profits, squeezed first by five years of tight labor markets and then by the initial impact of the recession, were in need of immediate relief; because its continuation would have led to a rate of unemployment so high that it could not possibly have been reduced to a politically acceptable level in time for the 1972 presidential election; and because the debt and liquidity structure of many corporations and banks raised questions as to whether they could survive the usual recessionary medicine.[15] As a result of recessed business conditions and less restrictive monetary policy in 1970, interest rates dropped by almost 60 percent by early 1971. This massive drop in *U.S.* short-term interest rates triggered equally massive short-term capital outflows. These outflows were combined with a trade balance which was being progressively eroded by inflation, demand-induced import growth, and Common Market tariff policies restricting *U.S.* agricultural exports. The outcome was a veritable explosion in the *U.S.* balance-of-payments deficit. By August 1971 this flood of *U.S.* dollars created such pressure on the international monetary system that an immediate response by the administration was essential.

---

[14] Evidence in support of this position is contained in a recent unpublished study by Thomas Riddell.

[15] That the financial condition of large corporations limits the use of restrictive monetary policy is understood by many knowledgeable observers of the economic scene. In a recent book, Leonard Silk, for example, asserts that "[the] liquidity squeeze . . . seemed to be threatening some businesses with financial disaster. . . . " He goes on to say: "The Federal Reserve's tight money policies during the latter part of 1969 were clearly involved in the liquidity squeeze. Chairman Burns . . . , facing a potential monetary panic, now threw Friedmanism to the winds and began to pour money into the economy. During the year following the near panic of May, 1970, M_1 grew by 8.5 percent and M_2 by 13 percent . . . " (p. 12).

B. *The Legacy of 1965–71*

The macropolicy decisions taken in support of the prosecution of the Vietnam War prolonged the American economic expansion through 1969—three or four years beyond its "natural" life. This extended expansion, culminating in five years of relatively full employment, had a number of significant effects on macrotheory and macropolicy.

It destroyed the credibility and general acceptance of the Keynesian promise of full employment, and it exposed the instability of the economic system. It eventually led to the imposition of government controls, signalling the end of the Keynesian "revolution." Whatever function macro-demand management may play in the future, the idea that the economy can be satisfactorily guided through Keynesian policies alone was permanently interred.

The domestic legacy of these years included a disproportionate buildup of personal and corporate debt, brought on in the corporate case by a sustained high level of investment and inventory spending undertaken in the context of declining profits and cash flow. Corporations were increasingly pushed to debt financing. In turn, the rise in indebtedness increased the tendency of the economy to develop liquidity crises and credit crunches. Put differently, it reduced the tolerance of the system for recession while increasing the necessity of a recession-induced reliquidation of the economy. In addition, the economy carried over both high rates of inflation and strong inflationary expectations.

The policies of the Vietnam period hastened the decline of the American monopoly of international capitalist power, consequently destroying the basis on which international stability rested. The unprecedented length of the *U.S.* expansion created an environment in which the export-oriented economies of Europe and Japan could sustain their own derivative expansions and accelerate their assault on our balance-of-trade surplus and our share of world export markets. The decade-long expansion in the world capitalist system in turn laid the foundation for a worldwide raw material or commodity inflation. In concert with the declining ability of the United States to control world events, the world commodity boom created the necessary preconditions for the coming economic and political revolt of the third world producers of raw materials, most significantly that of the oil-exporting countries. The balance-of-payments deficits and inflation-exportation of this period clearly hastened the demise of the existing international economic and financial institutions leaving little hope, in the power vacuum that existed, for a set of substitute institutions and arrangements.

In short, the expansive macropolicies chosen in the late 1960's to accommodate the defense of *U.S.* hegemony in the end accelerated its demise, leaving by August 1971 a legacy of international instability and an unmet need for a prolonged recession in a domestic system that seemed incapable of easily surviving one. It is thus clear why the administration was forced to abandon exclusive reliance on orthodox macropolicy. On the one hand, the system needed expansionary macropolicy to stimulate short-run profits and investment and to keep unemployment from rising too high. On the other hand, it needed restrictive policy to curb inflation, kick interest rates back up, improve the trade balance, and get labor fully under control. By August 1971 there was simply no combination of standard macropolicy tools capable of achieving these contradictory objectives. Nor could macropolicy deal with Europe's intransigence on the agricultural export question. The emergency in the international monetary system—a cornerstone of *U.S.* international hegemony—would not permit a delay in action. The administration's response was the New Economic Policy (*NEP*).

C. *The NEP and the Oil Crises: Setting the Stage for 1974*

We have argued that in order to explain the events of 1974 it is necessary to establish the position of the economy at the beginning of the year and evaluate the vector of forces affecting it. Specifically, the decision of the macropolicy makers to precipitate a semidepression is only comprehensible in light of such phenomena as rampant inflation, severe corporate financial stress, shortages amid massive inventory speculation, continued deterioration of the American international position, heightened international financial instastability, and the oil crisis. These forces posed the most serious threat to American capitalism since the 1930's. Analytically, then, the question at issue is why the economy was under more severe stress at the end of the *NEP* than it was in 1971 when the *NEP* was initiated.

The *NEP* was designed to achieve a set of objectives which were deemed contradictory in the absence of wage-price controls. Its proximate cause was the debacle in the international sector and, in our opinion, its chief objective was to rebuild the *U.S.* international position. Domestically, the *NEP* was intended to provide the framework within which orthodox macropolicy could generate an expansion which would both revive profits from their depressed level and prevent excessive inflation. The two objectives were obviously interrelated since control of domestic costs was essential to the rebuilding of international economic strength.

Most aspects of the international rationale for the *NEP* are clear. But the strategy of the "great agricultural export drive" has received little attention and merits discussion, particularly since the Report's chapter on food and agriculture fails to directly acknowledge the government's own contribution to food price inflation. The short- to medium-term strategy for protecting *U.S.* international interests was pretty well determined by the spring of 1971. A large trade surplus was considered essential for the financing of government spending and corporate investment abroad, and for providing a margin of safety against short-term capital flows. Secretary of the Treasury Connally stated that the trade balance had to move from a $4 billion deficit to a $9 billion surplus—a $13 billion turnaround. Agriculture was the only trade category which could be increased enough to cover a significant part of the requirement. In 1971 the Nixon Administration adopted an agricultural export policy to recoup its international losses.[16]

---

[16] The argument on the agricultural export drive referred to in this section is more fully developed in Boddy and Crotty (1975b). Their study analyzes the report and recommendations of the influential President's Commission on International Trade and Investment Policy. The Comission's report was formally presented to Nixon in July of 1971, although he was briefed on its contents at least as early as April of that year. The Commission's recommendations in support of the agricultural export drive and other international objectives later became the *NEP*. In the words of historian Joyce Kolko, the

The *NEP* was an integral part of this policy. On the demand side, the *NEP* contained two key provisions. The dollar devaluation was designed to make *U.S.* agricultural exports more competitive on world markets. The import surcharge represented an attempt to coerce Europe into negotiating a new agricultural trade policy to replace the Common Agricultural Policy (*CAP*), which had, since the middle 1960's, virtually excluded large scale American grain exports. The existence of an *EEC* import barrier in the context of a *U.S.* agricultural export drive explains the seemingly irrational decision of the administration to permit the massive sale of grains to the Soviet Union at subsidized prices.

The strategy on the supply side was somewhat more complex and intimately involved wage-price controls. By partially restraining price increases on those industrial goods which are agribusiness inputs, controls held down the cost of agricultural output. Farm products at first point of sale were specifically exempted from price controls. Agribusiness was thus assured of rising profit margins which, when combined with the removal of acreage restrictions, ensured a dramatic increase in production. Channelling this production into the export sector was accomplished in two ways. First, the combination of controlled wages and rising food prices meant that American workers could not afford their historic share of domestic agricultural production. Second, the lion's share of production would move to the export sector where prices were uncontrolled, while domestic food processors and retailers were subject to restrictions on the pass-through of cost increases. Higher profit margins could thus be obtained in the export sector.

The export drive was successful. The value of food exports in fiscal 1972 was $8.0 billion; by fiscal 1974 it had increased to $21.3 billion. Though the degree of success was dependent upon a series of crop failures around the world, a huge increase in agricultural exports was assured by the mechanics of the *NEP*, as was a maximal response to any unforeseen increase in world demand. The Report tells us that "every incremental increase in [agricultural] tonnage shipped . . . represents a corresponding reduction in the supply available to the domestic economy—and an increase in prices to the domestic economy" (pp. 183–84). One can only conclude that food price inflation was built into the *NEP* as a foreseeable and unavoidable consequence.

It is evident therefore that the *NEP* as a total package, including wage-price controls, was designed to deal with the short-term aspects of the international crisis largely via the "great agricultural export drive." In fact, there was no major facet of the *NEP* which did not support, directly or indirectly, the American effort to strengthen its international hand. In this context one can view the attempt to reassert control over domestic costs and prices as a dramatic statement to international competitors that the United States was determined to put its house in order. After years of deterioration the United States was attempting to regain its dominance in the world capitalist system, including the retention of reserve currency status for a strengthened dollar in a realigned fixed-rate system.

The manner in which the *NEP* handled the domestic problem of rising prices and shrinking profit margins through wage-price controls is also generally well understood, but several points relevant to an

---

Commission "advised a temporary import surcharge, a change of currency parities, wage and price controls, negotiations to eliminate trade barriers—in short, outlined Nixon's August 15, 1971 speech overturning the existing international monetary system" (p. 15).

explanation of economic conditions in 1974 need emphasis.

We agree with the Report's tentative conclusion that controls did not eliminate inflationary pressure, but rather postponed it from 1971–72 to 1974. Controls permitted the application of expansionary fiscal and monetary policy in 1972 and early 1973. This expansion would not have been possible without controls. The expansion in turn created pressures which the partial control system was incapable of withstanding. The inflationary thrust of the system remained when the control apparatus was disbanded in stages. There was a steep secular trend of inflation in early 1974 and a price-wage bulge following the removal of controls on April 30th of that year. Moreover, controls eventually led to surpluses and shortages because they suppressed market forces which, however socially irrational with respect to any reasonable list of national priorities, are a necessary tool of resource allocation in our system. One common source of domestic shortages was the increased allocation of various commodities to the export sector where profits were relatively high. And current shortages led to the fear of future shortages and thus to the kind of speculative and inflationary inventory buildups that plagued the economy in 1974. In addition, of course, the devaluations of the *NEP* contributed substantially to the rise in the rate of inflation.

Thus, controls did not in the end permit macropolicy to expand the economy in a noninflationary way. The rate of inflation in 1972 and 1973 was probably lower than it otherwise would have been given the macropolicy of this period. But the existence of controls was a necessary condition for the existence of very expansionary macropolicy so that it is misleading to take the policy stance as given.

Given the historical relation between real wages and the business cycle, the extended decline in real wages beginning in January 1973 is notable. From December 1972 to December 1973, the *CPI* rose by 8.8 percent while average hourly earnings in the private nonfarm sector were held to a 6.7 percent rate of growth. Real hourly earnings in the private nonagricultural sector, adjusted for interindustry employment shifts and in manufacturing for overtime, after rising 2.8 percent in 1972 (December to December), fell by 2.3 percent in 1973 and at a rate in excess of 5 percent in the first four months of 1974. Based on our reading of the evidence, it seems reasonable to conclude that the controls as applied in 1973 and early 1974 contributed to the decline in real wages and, consequently helped maintain the level and share of profits.[17] Tighter restrictions on real wages would have aborted the expansion earlier by further restricting consumption spending. Yet, the contribution of wage controls to the maintenance of profit and cash-flow margins was insufficient to avoid exacerbation of corporate debt and liquidity problems.

The expansion of 1972 and 1973 was accompanied and largely supported by an enormous expansion of corporate and consumer debt.[18] Nonfinancial and financial corporations entered 1974 with far worse liquidity and debt conditions than at the end of the expansion of the 1960's. In spite of the fact that profits in 1972–73 rose considerably from the 1970–71 lows, the financial structure of corporate America was in a precarious situation unprecedented in the post-World War II era. Internal sources of funds—though growing—fell increasingly short of investment and inventory accumulation in 1972 and especially 1973. An increasing proportion of corporate borrowing was short term, and corporate liquid assets as a percent of short-term debt declined precipitously. Common stock offerings of manufacturing companies declined to negligible levels by the second half of 1973. The *NEP* expansion was debt-fueled to a greater extent than any of its postwar predecessors.

---

[17] It has been argued by Franco Modigliani and others that real wages in 1973 were not restricted by the controls. The evidence used to support the argument is that the movement of nominal wages in this period ***given the actual price pattern*** was consistent with nominal wage equations estimated with pre-*NEP* data. We draw exactly the opposite conclusion from the same evidence. The dramatic rise of food prices and the *CPI* starting in January of 1973 must have affected actual price expectations almost instantaneously, whereas the wage equations of pre-1971 vintage assume that a substantial time lag exists between a change in the price level and the incorporation of that change into price expectations and thus into nominal wage changes. That nominal wages adjusted as slowly to the galloping inflation of 1973 and early 1974 as they did to the creeping inflation of the 1960's therefore would seem to be evidence in favor of the proposition that controls retarded the growth of real wages. The fact that the decline in real wages was temporarily reversed in the months immediately following the termination of controls lends additional support to this proposition.

The *NEP* was an innovative policy undertaken by a capitalist government to achieve the complementary objectives of regaining America's international hegemony and restoring "balance" to the domestic economy. The mechanism was to be an economic expansion immunized against inflation and a profit squeeze through wage and price controls. It basically failed on both fronts. The *NEP* carried over unresolved into 1974 many of the domestic and international problems facing the *U.S.* economy in August 1971, and it severely exacerbated or created several others: inflationary pressures in general, and food price inflation in particular; debt and liquidity problems; shortages, speculative inventory buildups, and commodity inflation; secular profit and cashflow problems, temporarily alleviated by the controlled decline in real wages; the domestic inflationary impact of the declining value of the dollar; a fundamental, unresolved instability in the international sector. The basic problems of the economy in 1971 remained as 1974 began. The economy suffered from severe imbalances and disproportions requiring a massive recession it might not be able to withstand, and was situated in an international context of declining *U.S.* power amid system instability.

Some economists deny the assertion that the present international financial arrangements are unstable. They believe that floating exchange rates are the only mechanism consistent with international economic order. But we believe that the system of floating rates which emerged by default in the 1970's cannot serve as the basis for international economic stability. Floating rates are an interim arrangement that reflects the breakdown of international order attendant upon the relative decline in American power. It is understandable, therefore, that the Report's tone in its international section is one of relief that the system survived the year 1974.

The Report's discussion of floating rates is descriptive and uncritical. No attempt is made to celebrate the virtues of floating as opposed to fixed rates. No detailed discussion or analysis is made of the main issues traditionally associated with the flexible-fixed exchange rate discussion—independence of monetary-fiscal policy, exchange rate stability, depreciation and inflation, and trade deficit elimination through depreciation. The absence of discussion of these issues leaves the reader with the impression that the Report's authors understand that until a political resolution is reached, the world capitalist system will at best limp along with its managed float. But the Report underestimates the risks inherent in the present system. Widespread depression, oil import deficits, and the absence of political agreement are good reasons to fear protectionist policies and a major collapse in world trade.[19]

---

[18] See the special issue on corporate and consumer debt of *Business Week*, Oct. 12, 1974.

The proponents of floating rates, who have offered what are now seen by many as false promises, have met with widespread criticism.[20] Floating rates neither automatically equilibrate trade balances nor create independence of action for the monetary-fiscal authorities. Unless macropolicies are coordinated across national boundaries, policy actions trigger interest rate differentials which—in the context of enormous amounts of interest sensitive short-term capital, thin foreign exchange markets, and weak or uncertain expectations concerning future rates—exert effective and continuous pressure on exchange rates. Since policy makers are intensely concerned about the international position of their currency, the degree of policy independence created by managed floating is quite small.[21] Exchange rates, moreover, have been quite volatile since March of 1973. During 1974, for example, the dollar first fell 17.5 percent against the Deutschemark, then rose by 9.5 percent only to fall again to an end of the year low. Finally, in the absence of downward domestic price flexibility, dollar devaluation is a source of domestic inflationary pressure and exchange rate fluctuations are a source of world inflationary pressures.

In the long period of Pax Britannica and in the short period of Pax Americana, a fixed exchange rate system was the choice of the dominant economic power. Like any monopolist, the dominant power preferred to set prices rather than subject itself to the vagaries of the market. The existence of the current floating rate system is, in other words, proof that no international power is presently capable of reconstituting a new stable fixed-rate regime. While currently unavoidable, the present system is itself a contributor to international uncertainty and instability.

By late 1973 the macropolicy authorities were faced not only with all of the domestic and international problems and contradictions left unresolved by the *NEP*, but were confronted with an oil crisis as well. The Report devotes considerable attention to the oil problem, but fails to discuss the critical role played by the decline in *U.S.* world power in the development of this crisis. So long as this country enjoyed a virtual monopoly on world economic, political, and military power, it could depend on a guaranteed supply of cheap imported energy and raw materials. A weakened international position, however, made the United States unable to prevent the oil-producing countries from quadrupling the price of oil. That could not have happened in the 1950's—witness the *U.S.* orchestrated overthrow of the Iranian Prime Minister, Mossadegh, in 1953—and can therefore be taken as an extreme example of the process by which international instability destroys the basis for domestic stability.

IV. The Recession of 1974–75

Except for the introduction of extensive economic planning, the administration had no option other than to permit the economy to recess in late 1973. The stakes were high. The accumulated pressures described in the preceeding sections were more severe in 1973 than in late 1969. The controls were under extreme market pressure; the financial structure of businesses was imbalanced because of extensive debt financing; the rate of inflation was accelerating; and the international economy was in disarray. Nothing short of a major recession-depression could eliminate these imbalances from the system. While the impact of recessionary policies was uncertain, these policies were viewed as less risky than any of the perceived alternatives. The economy had to be purged. First fiscal and then monetary policy accomodated to this need.[22] Fiscal policy was restrictive beginning in the middle of 1973 and remained so throughout 1974. Monetary policy tightened in the second half of 1973 and became severely restrictive in the second half of 1974. The macropolicy authorities were attempting to generate a severe economic contraction. Skyrocketing unemployment rates in late 1974 left no doubt of their success. This policy jeopardizes a Republican victory in the 1976 elections just as the Democrats' Vietnam policy undermined their 1968 election prospects. But in both cases the policies were essential in order for the administrations to fulfill their responsibility to defend capitalist institutions. Unemployment and stagnation were the chosen first lines of defense.

---

[19] The possibility that a world depression might result from the relative decline in *U.S.* world hegemony is stressed by Kindleberger in his book on the Great Depression. He expresses the hope that either "cession of economic sovereignty to international institutions" or revival of American leadership will avoid such a crisis. For a more pessimistic outlook for world capitalism in the face of interimperialist rivalry between the EEC countries and the United States, see the Marxist analysis of Ernest Mandel.

[20] For a balanced review of the experience under floating rates, see Marina Whitman and comments by Paul Volcker. For a stronger dose of criticism of floating rates, see the comments of Kindleburger, Arthur Laffer, Paul Samuelson, and Volcker in *Business Week*, June 2, 1975.

[21] That the effect of monetary policy on the international value of the dollar constrains the domestic monetary authorities is certainly made clear in the public statements of Chairman Burns. For example, Edwin Dale in a recent New York Times column comments: "Arthur F. Burns . . . is letting it be known that a major constraint on how much the Federal Reserve can do is the international exchange rate of the dollar."

V. Prospects for the Future

Except for some tentative projections on unemployment and inflation rates in 1975, the Report itself does not speculate on the prospects for the remainder of this decade. The Office of Management and Budget has, however, provided the administration's projections for this period, and they present a dismal picture for working people. Starting from the optimistic assumption that the economy will witness an uninterrupted growth in real *GNP* averaging about 6 percent a year from 1976 through 1980, they project an average reported unemployment rate for the period of almost 7 percent, making the 1970's a decade of joblessness. To perceive the human impact of these forecasts, consider that the fraction of the work force who experience unemployment at some time during the year may be three or more times the average unemployment rate.

These projections, however, are not accompanied by any serious political-economic analysis of the forces likely to be operating on the *U.S.* economy over the remainder of the decade. Yet powerful secular pressures will clearly influence policy and performance. There will be a drive to reassert American control over a restabilized international economy, to reverse the decade-long decline in *U.S.* productivity, to close the gap between the percent of output channelled into capital accumulation here relative to Europe and Japan, to obtain what the Report calls "energy security," and to increase the rate of profit and decrease the growth of corporate debt. These objectives will not only influence policy but may lead to class struggle at home and rivalry and conflict abroad. They may also, in concert with the generally abysmal performance of the system over the past decade, induce major institutional change.

Given the conjunction of these pressures with the current economic stagnation, every option confronting those who control economic policy involves economic contradictions and policy conflict. American capital must engineer a long, noninflationary, balanced economic expansion within which the rate of capital accumulation dominates consumption growth. There appears to exist no single policy, and certainly no monetary-fiscal mix, that can achieve these objectives.

---

[22] In 1972, the high-employment budget was $7.6 billion in deficit. By the third quarter of 1973, this budget was in surplus and remained so throughout 1974. At an annual rate, the budget surplus was $8.1 billion in the fourth quarter of 1973 and $11.5 billion over the year 1974. The annual growth of the nominal money stock (M_1) was 8.6 percent from the beginning of 1972 through mid-1973. It declined to 5.5 percent over the next year. Then it fell precipitiously to 1.4 percent in the last half of 1974. Because of accelerating inflation in 1973–74, these figures severely understate the restrictiveness of monetary policy.

One option would be to initiate a strongly expansionary macropolicy which would generate full employment rather quickly. But this would lead to rapidly accelerating inflation, something neither the domestic nor the international economy can tolerate. A rapid *U.S.* expansion would lead to an international economic revival retriggering commodity inflation and a large increase in the world demand for oil—an ominous prospect indeed. Too many structural imbalances remain in the domestic economy to make possible a repeat of 1961 to 1965.

Another option would be to generate a long, gradual expansion much like the one projected in the budget in the hope that the years of high unemployment would hold wages down while production slowly edged upward. Meanwhile, the drive for a discrete jump in the rate of capital accumulation would be aided by fundamental changes in the tax system. Capital's demand for corporate tax relief has been widely supported. Every week brings new proposals by economists, politicians, or influential individual or organizational business spokesmen to lower the effective rate of corporate taxation and/or the income tax rate of the wealthy. A further move in the direction of a higher investment-consumption ratio may also be discerned in the current fiscal crises of state and local governments. In the context of budget crises, state and local governments are in the process of purging their welfare and employment rolls. The federal government apparently does not desire to relieve these fiscal pressures. This process, should it continue, will transfer resources from the public to the private sector and, by adding to the reserve army of the unemployed, will increase the downward pressure on wage rates.

The gradual expansion option has contradictions. Not only would it fail to guarantee price stability, but its long years of intolerable unemployment rates raise the prospect of heightened class conflict in the United States. Although the leaders of organized labor have been incredibly docile up to this point, they may not be able to restrain rank and file anger for the duration of the decade. A slow recovery would create even more acute labor problems in Europe and perhaps even Japan. Finally, the desired acceleration of capital accumulation might not be sustainable if working class incomes and consumption demand are restricted. Even a gradual expansion may abort under these circumstances.

Alternatively, the government might initiate an incomes policy of wage-price and perhaps profit and interest rate controls administered in a manner similar to those of the *NEP*. These controls might be used to augment one of the monetary-fiscal strategies discussed above. A target rate of inflation, coupled with a rule that money wage rate changes must on average equal this target rate of inflation plus the rate of growth of productivity, might set the stage for expansive macropolicy by freezing labor's share. But if we add an unusually high target rate of capital accumulation, a reduction of labor's share would be necessary.

Such an augmented incomes policy poses difficult questions. Can demand be sustained in the face of a declining share of income for labor? Will labor tolerate a lower share of output? An incomes policy, moreover, is essentially self-contradictory in the context of a capitalist economy. We accept as generalizable the Report's assertion that the *NEP* at some point in the expansion "began to obstruct normal supplier-purchaser relationships, and in some cases the controls became quite unworkable" (pp. 228–29). Unless more extensive micro or sectoral planning is introduced, a permanent system of wage-price controls is untenable.

Economic planning may be the wave of the future, even under capitalism; it represents an option, perhaps *the* option in the long run. Several notable planning proposals have already been made but they do not envision a fully planned economy as we normally understand this concept. They deal more with gathering information, making forecasts, using some tax and subsidy incentives and marginally altering resource allocation. But it is clear to us that if planning is to help resolve major contradictions in the domestic economy, it must achieve a significant substitution of administrative for market processes. Genuine planning could involve not only the setting of wages and prices, but government-directed allocation of raw materials and credit, a detailed system of tax subsidies and penalties, antistrike or even anticollective bargaining legislation, and administratively coordinated investment strategies among firms and industries.

Although pressure toward planning is likely to continue, it will not be steady. While a brief period of relatively trouble-free expansion may temporarily silence supporters of planning, renewed outbreaks of inflation, rising unemployment, credit crises, and/or shortages will intensify this pressure. Planning may represent the only way to eliminate recurrent crises from the domestic economic system and to successfully pursue the goal of *U.S.* domination of the world capitalist system. Only within a planned economy is it potentially possible to achieve full employment and price stability and at the same time restrain wages and consumption in support of capital accumulation. But planning demystifies economic power relations and transfers the stresses and strife of the market system into the political arena. It therefore portends political conflict of the highest order—conflict within the capitalist class as well as between capital and labor.

Within the present array of political forces in the United States, the development of economic planning will undoubtedly mean corporate control of the planning process, just as the introduction of federal regulatory agencies has historically meant control by and for the regulated industries. It is, therefore, more important than ever before that the political balance of power be changed and the power of the corporate elite broken. Working people and their allies must defend the interests of the majority of the American people against the fundamentally antagonistic interests of the corporations and the super-rich who own and control them.

REFERENCES

R. Boddy and J. Crotty, (1975a) "Class Conflict and Macro-Policy: The Political Business Cycle," *Rev. Rad. Polit. Econ.*, Spring 1975, *7*, 1–19.

——— and ———, (1975b) "Food Prices: Planned Crisis in Defense of the Empire," *Socialist Revolution*, Apr. 1975, *23*, 101–11.

S. Bowles and H. Gintis, "The Problem with Human Capital Theory—A Marxian Critique," *Amer. Econ. Rev. Proc.*, May 1975, *65*, 74–82.

A. Burger, "Relative Movements in Wages and Profits," *Fed. Reserve Bank St. Louis Rev.*, Feb. 1973, *55*, 8–16.

J. Crotty, "The Objective Function of the State: A Study of the Political Economy of Monetary and Fiscal Policy," unpublished doctoral dissertation, Carnegie-Mellon Univ. 1973.

J. Gurley, "Have Fiscal and Monetary Policies Failed?," *Amer. Econ. Rev. Proc.*, May 1972, *62*, 19–23.

D. Hibbs, "Industrial Conflict in Advanced Industrial Societies," Center Internat. Stud., Cambridge, Apr. 1974.

T. Hultgren, *Costs, Prices, and Profits: Their Cyclical Relations,* New York 1965.

M. Kalecki, "Political Aspects of Full Employment," in his *Selected Essays on the Dynamics of the Capitalist Economy*, Cambridge 1971, 138–45.

J. M. Keynes, *The General Theory of Employment, Interest, and Money,* New York 1964.

C. P. Kindleberger, *The World in Depression: 1929–1939,* Berkeley 1973.

J. Kolko, *America and the Crisis of World Capitalism,* Boston 1974.

M. Landsberg, "Black Gold and the Reemergence of the Dollar Standard," *Rev. Rad. Polit. Econ.*, Spring 1975, *7*, 69–78.

A. MacEwan, "The Current Economic Crisis," *Radical Amer.*, No. 1, *9*, 1–23.

E. Mandel, *The Decline of the Dollar: A Marxist View of the Monetary Crisis,* New York 1972.

K. Marx, *Capital,* Vol. 1, New York 1967.

W. C. Mitchell, *Business Cycles and Their Causes,* Berkeley 1941.

F. Modigliani, "The 1974 Report of the President's Council of Economic Advisors: A Critique of Past and Prospective Economic Policies," *Amer. Econ. Rev.*, Sept. 1974, *64,* 544–57.

G. H. Moore, "Productivity, Costs and Prices: New Light From an Old Hypothesis," unpublished paper, Apr. 28, 1974.

W. Nordhaus, "The Falling Share of Profits," *Brookings Papers,* Washington 1974, *1,* 169–208.

A. Okun, *The Political Economy of Prosperity,* New York 1970.

T. Riddell, "A Political Economy of the American War in Indo-China: Its Costs and Consequences," unpublished doctoral dissertation, American Univ. 1975.

L. Silk, *Nixonomics,* New York 1972.

P. M. Sweezy and H. Magdoff, *The Dynamics of U.S. Capitalism: Corporate Structure, Inflation, Credit, Gold, and the Dollar,* New York 1972.

H. Stein, *The Fiscal Revolution in American,* Chicago 1965.

P. A. Volcker, "Domestic Adjustment to Balance of Payments Disequilibrium: Discussion," *Amer. Econ. Rev. Proc.*, May 1975, *65,* 153–56.

M. v.N. Whitman, "The Payments Adjustment Process and the Exchange Rate Regime: What Have We Learned?," *Amer. Econ. Rev. Proc.*, May 1975, *65,* 133–46.

Business Week, Oct. 12, 1974; June 2, 1975.

U.S. Council of Economic Advisers, *Economic Report of the President,* Washington 1975.

Securities Market Efficiency in an Arrow-Debreu Economy

By MARK RUBINSTEIN*

The word "efficiency" as applied to securities markets has unfortunately been used to represent a variety of logically distinct concepts. In particular, it may mean:

> a) Exchange efficiency: participants are not motivated to create exchange arrangements not already provided by the market.
>
> b) Production efficiency: value-maximizing firms make Pareto optimal production decisions.
>
> c) Information efficiency: present security prices are costlessly known to all participants and information about future security prices is "fully reflected" in these present prices.

To focus undiluted attention on information efficiency, the concepts a) and b) will be automatically satisfied by the Arrow-Debreu model used in this paper.[1] Moreover, since transactions costs are assumed nonexistent (i.e., perfect market), the remaining question and chief focus of this paper is: when will present security prices fully reflect information about future prices?

This question should first serve to raise the logically prior question, what do we mean by the phrase "fully reflect information"? In a perfect and competitive economy composed of rational individuals with homogeneous beliefs about future prices, by any meaningful definition present security prices must fully reflect all available information about future prices. According to Eugene Fama,

> First, it is easy to determine *sufficient* conditions for capital market efficiency. For example, consider a market in which (i) there are no transactions costs in trading securities, (ii) all available information is costlessly available to all market participants, and (iii) all agree on the implications of current information for the current price and distributions of future prices of each security. In such a market, the current price of a security obviously "fully reflects" all available information. [p. 387]

Moreover, the model of the securities market developed here will be informationally efficient in this trivial case. The conditions for efficiency only become interesting when participants have heterogeneous beliefs (or face positive transactions costs, not considered here). In this case, it seems obvious that some may perceive their information as fully reflected in present prices, while others may not. The challenge is to identify who is who, or to put it somewhat more abstractly, to identify what information is and is not reflected in present prices. A precise analytical definition of information-efficiency is developed below. I shall argue that despite some attempts (see Fama and Paul Samuelson), its definition under nontrivial circumstances has remained ambiguous.

---

* Assistant professor, Graduate School of Business Administration, University of California, Berkeley. Research for this paper was supported in part by a grant from the Dean Witter Foundation. I am indebted to Jack Hirshleifer for many discussions as well as previous collaboration. However, I alone bear any responsibility for errors in this paper.

[1] In the context of an incomplete market for securities, exchange and production efficiency both hold whenever there is universal portfolio separation so that at each date all individuals divide their wealth among consumption and the same two mutual funds, one of which is risk free. (See the author, 1973, 1974.)

Section I details a model constructed to answer these questions. Section II develops a precise definition of information efficiency and identifies the characteristics of participants who believe present security prices fully reflect their information. Section III examines implications for the intertemporal structure of security rates of return (random walk, unbiased term structure). Section IV contrasts the definition of information efficiency with those of Fama and Samuelson. Section V mentions the chief limitations of the model in view of the issues addressed.

I. The Economy

Appropriate answers to the questions posed require a model of a dynamic, uncertain, heterogeneous[2] economy. Assume each individual in the economy makes decisions over a three-date (t=0, 1, 2) horizon. At date t=0, he chooses present consumption C_0 and makes a *provisional* choice of future consumption by selecting a portfolio of contingent claims $\{\hat{C}_e\}$ to date t=1 consumption which pay off only if state *e* occurs at date t=1, and by selecting a portfolio of contingent claims $\{\hat{C}_s\}$ to date t=2 consumption which pay off only if state *s* occurs at date t=2. Consequently, if $\{P_e\}$ and $\{P_s\}$ denote, respectively, the date t=0 prices to these claims,[3] the individual will divide his date t=0 wealth W_0 so that $C_0+\Sigma_e P_e\hat{C}_e+\Sigma_s P_s\hat{C}_s=W_0$.

At date t=1, the true state *e* occurs: contingent claims to that state pay off, contingent claims to other states at that date become worthless, and new information becomes available and is dispersed perhaps unevenly throughout the economy. This information revises beliefs held by individuals concerning prices to rule in the market at date t=2. As a consequence, and also due to the mere passage of time, the equilibrium prices of contingent claims to date t=2 consumption are revised from their prior levels set at date t=0. Since these revised prices will in general depend on state *e*, they are denoted $\{P_{s\cdot e}\}$. The individual's wealth W_e at date t=1 will therefore depend both on state *e* and on his prior choices so that $W_e\equiv\hat{C}_e+\Sigma_s P_{s\cdot e}\hat{C}_s$. In view of the revised prices, each individual with his own revised beliefs will in general desire to revise his provisional choices $\hat{C}_e$ and $\{\hat{C}_s\}$ made at date t=0. To accommodate this, denote C_e and $\{C_{s\cdot e}\}$ as his revised (and final) choice of date t=1 consumption and portfolio of contingent claims to date t=2 consumption (in general, $C_e \neq \hat{C}_e$ and $\{C_{s\cdot e} \neq \hat{C}_s\}$). The subscript e reflects that his revision at date t=1 will in general depend on the true state at that date and consequently, after revision, to satisfy his wealth constraint $C_e + \Sigma_s P_{s\cdot e} C_{s\cdot e} = W_e$. Date t=1 is the last time choices are made. At date t=2 the true state s occurs: contingent claims to that state pay off, and contingent claims to other states become worthless. Consequently, the individual simply consumes $C_{s\cdot e}$.

---

[2] Specifically, resources, beliefs, and tastes are generally different for different individuals while they all share the same opportunities (i.e., face the same prices of securities—an implication of competitive markets).

[3] A type of complete market has been assumed. For complete markets in the Debreu sense (pp. 98–102), a complete set of consumption claims to date t=2 consumption would need to be available at date t=0 with payoffs if and only if *both state s and state e* occur. With these claims available, the exchange opportunities at date t=0 would be sufficiently rich that no individual would need to revise his portfolio at date t=1 and a market at that date would be redundant. It is well known that if such claims exist, then the equilibrium will be Pareto optimal; in particular, no individual will have an incentive to create additional securities or engage in transactions outside the market. However, these claims do *not* exist in the economy described in this paper; nonetheless, the equilibrium will still be Pareto optimal since it can be shown to yield exactly the same allocations as if these claims did exist. The general principle is that individuals are indifferent between relatively rich forward markets with few future revision opportunities and relatively poorly developed forward markets with many future trading dates. In reality, market setup costs and economies of scale in exchange will help determine this tradeoff. The objections raised against the assumption of rich forward markets lose a great deal of force if one recognizes that individuals may be able to make the same final choices even in the absence of some forward markets, provided sufficient opportunities for revision exist. One need only recall the virtually continuous trading provided by stock exchanges to lend strength to this position. See Kenneth Arrow and commentary by Jacques Dreze, pp. 144–45, for the genesis of this argument. In particular, it presumes future security prices are correctly forecast; see Section V of this paper for further discussion.

All individuals are presumed to obey the Savage axioms of rational choice: they have heterogeneous beliefs and tastes, representable by probabilities and utility functions, so that they make choices which maximize expected utility. The term $\pi_e > 0$ denotes the probability held by an individual that state e will occur; given the occurrence of state e, $\pi_{s\cdot e} > 0$ denotes the conditional probability that state s will occur; by the laws of probability, the unconditional probability that state s will occur $\pi_s = \Sigma_e \pi_e \pi_{s\cdot e}$. Just as prices of contingent claims to date t=2 consumption may be revised at date t=1, beliefs may be similarly revised (in general, $\{\pi_{s\cdot e} \neq \pi_s\}$). Each individual is assumed to maximize an additive, state-independent utility function over his final consumption plan (C_0, $\{C_e\}$, $\{C_s\}$). That is, the utility of consumption $U_t(\cdot)$ at date t is evaluated independently of consumption at all other dates (but may depend on the date itself) and independent of states.[4]

The following programming problem summarizes the above description:[5]

$$\max_{C_0, \{\hat{C}_e\}, \{\hat{C}_s\}, \{C_e\}, \{C_{s\cdot e}\}} \quad U_0(C_0) + \Sigma_e \pi_e U_1(C_e)$$
$$+ \Sigma_e \pi_e \Sigma_s \pi_{s\cdot e} U_2(C_{s\cdot e})$$

subject to $W_0 = C_0 + \Sigma_e P_e \hat{C}_e + \Sigma_s P_s \hat{C}_s$

$$C_e + \Sigma_s P_{s\cdot e} C_{s\cdot e} = W_e \equiv \hat{C}_e + \Sigma_s P_{s\cdot e} \hat{C}_s \qquad \text{(all } e\text{)}$$

Factoring and using Lagrangian multipliers,[6] the problem may be restated

$$\max_{C_0, \{\hat{C}_e\}, \{\hat{C}_s\}, \{C_e\}, \{C_{s\cdot e}\}} \quad U_0(C_0)$$
$$+ \Sigma_e \pi_e U_1(C_e) + \Sigma_s \pi_{s\cdot e} U_2(C_{s\cdot e})$$
$$- \lambda_e[C_e + \Sigma_s P_{s\cdot e} C_{s\cdot e} - \hat{C}_e - \Sigma_s P_{s\cdot e} \hat{C}_s]$$
$$- \lambda_0[C_0 + \Sigma_e P_e \hat{C}_e + \Sigma_s P_s \hat{C}_s - W_0]$$

The necessary and sufficient conditions for an optimum are determined by differentiating the maximand partially with respect to each choice variable and repeating the constraints. Differentiating successively by C_0, $\{\hat{C}_e\}$, $\{\hat{C}_s\}$, $\{C_e\}$, and $\{C_{s\cdot e}\}$,

$$U'_0(C^*_0) = \lambda_0$$

---

[4] For a more general portfolio model, excepting its assumption of constant proportional risk aversion, see Nils Hakansson, who allows for incomplete markets, uncertain lifetime, and state-dependent utility. The salient feature of his model, as well as the one presented in this paper, is the freedom of prices to follow the most generalized (discrete-time) stochastic process.

[5] Assume that $U'_t > 0$ and $U''_t < 0$. Nonnegativity constraints on actual consumption (C_0, $\{C_e\}$, $\{C_{s\cdot e}\}$) are to be satisfied although they are not explicitly incorporated into the analysis. For at least one class of utility functions, U_t, those with constant proportional risk aversion, the nonnegativity constraints are automatically satisfied since for this class the marginal utility of consumption is infinite at zero consumption.

[6] Use here has been made of stochastic Lagrangian multipliers, which are a natural generalization of standard Lagrangian techniques. (For example, see Dreze.) Formulating the programming problem in this way permits it to be solved without explicitly working backwards via a derived utility function as in dynamic programming. Had we done so, for the date t=1 subproblem contained in the large brackets, the derived utility of wealth function

$$V_e(W_e) \equiv \max_{C_e, \{C_{s\cdot e}\}} U_1(C_e) + \Sigma_s \pi_{s\cdot e} U_2(C_{s\cdot e})$$
$$- \lambda_e[C_e + \Sigma_s P_{s\cdot e} C_{s\cdot e} - W_e]$$

where $W_e = \hat{C}_e + \Sigma_s P_{s\cdot e} \hat{C}_s$. In general, even though utility over consumption is state-independent, the derived utility function $V_e(W_e)$ of date t=1 wealth will not be state independent. It will depend on state e not merely indirectly through W_e but also through the revised state-dependent prices and beliefs. With this subproblem solved, the date t=0 problem reduces to

$$\max_{C_0, \{\hat{C}_e\}, \{\hat{C}_s\}} U_0(C_0) + \Sigma_e \pi_e V_e(W_e)$$
$$- \lambda_0[C_0 + \Sigma_e P_e \hat{C}_e + \Sigma_s P_s \hat{C}_s - W_0$$

$$\pi_e\lambda_e = \lambda_0 P_e \qquad \text{(all } e\text{)}$$

$$\Sigma_e \pi_e \lambda_e P_{s\cdot e} = \lambda_0 P_s \qquad \text{(all } s\text{)}$$

$$U_1'(C_e^*) = \lambda_e \qquad \text{(all } e\text{)}$$

$$\pi_{s\cdot e} U_2'(C_{s\cdot e}^*) = \lambda_e P_{s\cdot e} \qquad \text{(all } s \text{ and } e\text{)}$$

The stochastic Lagrangian multipliers $\{\lambda_e\}$ play the role of linking the decision problems at dates t=0 and t=1. Eliminating the Lagrangian multipliers and rearranging, the necessary and sufficient conditions for an optimum can be symmetrically represented by

$$(1) \qquad P_s = \Sigma_e P_e P_{s\cdot e} \qquad \text{(all } s\text{)}$$

$$(2) \qquad \pi_e U_1'(C_e^*) = U_0'(C_0^*) P_e \qquad \text{(all } e\text{)}$$

$$(3) \qquad \pi_{s\cdot e} U_2'(C_{s\cdot e}^*) = U_1'(C_e^*) P_{s\cdot e} \qquad \text{(all } e \text{ and } s\text{)}$$

$$(4) \qquad W_0 = \Sigma_s P_s \hat{C}_s^* + \Sigma_e P_e \hat{C}_e^* + C_0^*$$

$$(5) \qquad \Sigma_s P_{s\cdot e} C_{s\cdot e}^* + C_e^* = \Sigma_s P_{s\cdot e} \hat{C}_s^* + \hat{C}_e^* \quad \text{(all } e\text{)}$$

Condition (1) is an arbitrage relationship among prices, paralleling $\pi_s = \Sigma_e \pi_e \pi_{s\cdot e}$, which must hold in equilibrium. Conditions (2) and (3) are the standard single-period optimality conditions linking marginal utility to prices, and conditions (4) and (5) repeat the constraints.

Let i superscript individuals and ($\overline{C}_0^i$, $\{\overline{C}_e^i\}$, $\{\overline{C}_s^i\}$) represent the endowment of individual i so that $W_0^i \equiv \overline{C}_0^i + \Sigma_e P_e \overline{C}_e^i + \Sigma_s P_s \overline{C}_s^i$. The closure conditions then needed to endogenously determine equilibrium prices would be

$$\Sigma_i C_0^i = \Sigma_i \overline{C}_0^i$$

$$\Sigma_i C_e^i = \Sigma_i \hat{C}_e^i = \Sigma_i \overline{C}_e^i \qquad \text{(all } e\text{)}$$

$$\Sigma_i C_{s\cdot e}^i = \Sigma_i \hat{C}_s^i = \Sigma_i \overline{C}_s^i \qquad \text{(all } e \text{ and } s\text{)}$$

The i superscript will continue to be omitted, unless there is danger of ambiguity.

II. Nonspeculation and Consensus

We will now use conditions (1)–(5) to determine what information about future prices is fully reflected in present prices. We shall first provide a precise definition of prices that fully reflect the *new information* that becomes available to an individual at date t=1; then this will be extended to a definition of prices that fully reflect *all information* available to an individual at both dates t=0 and t=1.

An individual will be said to perceive the new information that becomes available to him as fully reflected in revised security prices if and only if he has "nonspeculative beliefs." By definition, *nonspeculative beliefs are those beliefs for which no portfolio revision is an optimal strategy.*[7] The model described in Section I has been constructed to motivate this definition.

Observe that the variety of consumption claims available at date t=0 is just sufficient for an individual to move directly to his optimal final consumption plan by choosing $\{C_e^* = \hat{C}_e^*\}$ and $\{C_{s\cdot e}^* = \hat{C}_s^*\}$. In this case, he opts not to revise his portfolio at date t=1, but stays with his date t=0 provisional choices. This might well be optimal since one of the chief reasons for portfolio revision—availability of new consumption claims at future dates—is absent. At date t=0, an individual can purchase claims to consumption at every state that might occur at dates t=1 and t=2. Other possible causes of portfolio revision—barriers to marketability occasioned by nonmarketable resources, short selling restrictions, and finite transactions costs, as well as state-dependent tastes—are omitted from the model.

---

[7] Joseph Stiglitz also identifies speculation with portfolio revision. Using a three-date model, he examines the relationship between the term structure and an individual's demand for short- vs. long-term bonds. At date t=0, individuals choose between short- and long-term bonds. At date t=1, the previous short-term bonds mature and the previous long-term bonds now become the only existing short-term bonds. Individuals can revise their portfolios, choosing between consumption and the short-term bonds. At date t=2, they consume their remaining wealth. Stiglitz introduces uncertainty by assuming the date t=1 price of the long-term bonds is uncertain at date t=0. Unfortunately, since Stiglitz admits no purely risky securities into his model, in an equilibrium context his economy is riskless and the injected uncertainty disappears together with the purpose of his model.

In fact, the economy has been so constructed as to leave only one reason for portfolio revision: a change in prices to claims to date t=2 consumption that is not fully offset by revised beliefs. In general, for claims under some state s, price changes from P_s at date t=0 to $P_{s\cdot e}$ at date t=1 if state e occurs. Similarly, on the basis of newly received information about the future state s, an individual revises his belief from π_s at date t=0 to $\pi_{s\cdot e}$ at date t=1 if state e occurs. Since his portfolio chosen at date t=0 was based on price P_s and belief π_s, with revised price and belief at date t=1, the individual will generally desire to revise his portfolio. For example, even without formal analysis, it should be intuitively obvious that if $P_{s\cdot e}<P_s$ and $\pi_{s\cdot e}>\pi_s$, then the individual will typically shift more of his holdings into consumption claims to state s (i.e., $C^*_{s\cdot e}>\hat{C}^*_s$).

However, if the change in price from P_s to $P_{s\cdot e}$ were in some sense precisely offset by the change in belief from π_s to $\pi_{s\cdot e}$ and this occurred for all states s and e, then the individual would plan not to revise his portfolio at date t=1 regardless of the state e that occurs at that date. If this offsetting effect occurs, then we shall say the individual has nonspeculative beliefs and revised prices fully reflect the new information received by the individual at date t=1. Consequently, to identify nonspeculative beliefs and prices that fully reflect new information, we need only examine the circumstances under which no portfolio revision is optimal.

THEOREM (Nonspeculation Condition): *No portfolio revision under all states e is optimal* ($\{C^*_e=\hat{C}^*_e\}$ *and* $\{C^*_{s\cdot e}=\hat{C}^*_s\}$) *if and only if*

$$\frac{\pi_s}{P_s}=\frac{\pi_e\pi_{s\cdot e}}{P_eP_{s\cdot e}}\qquad (\textit{all } s \textit{ and } e)$$

PROOF:

Applying conditions (1)–(5), we will (i) develop necessary and sufficient conditions for $C^*_{s\cdot e}$ to be independent of e; and (ii) show that in this case, given $C^*_{s\cdot e}$ and C^*_e, it is feasible to move immediately to these choices at date t=0.

(i) Substituting condition (2) into (3):

$$\frac{U_0'(C_0^*)}{U_2'(C^*_{s\cdot e})}=\frac{\pi_e\pi_{s\cdot e}}{P_eP_{s\cdot e}}$$

If $C^*_{s\cdot e}$ is independent of e, then there exists a number k_s independent of e so that $k_s\equiv(\pi_e\pi_{s\cdot e})/(P_eP_{s\cdot e})$. Summing over states e, $k_s(\Sigma_eP_eP_{s\cdot e})=\Sigma_e\pi_e\pi_{s\cdot e}$, which by condition (1) and the laws of probability implies $k_sP_s=\pi_s$. Consequently, $\pi_s/P_s=(\pi_e\pi_{s\cdot e})/(P_eP_{s\cdot e})$. Moreover, if this last equation holds, then $U_0'(C_0^*)/U_2'(C^*_{s\cdot e})=\pi_s/P_s$ so that $C^*_{s\cdot e}$ must be independent of e. In summary, we can define C^*_s so that $C^*_{s\cdot e}\equiv C^*_s$ if and only if $\pi_s/P_s=(\pi_e\pi_{s\cdot e})/(P_eP_{s\cdot e})$. This holds for all s and e.

(ii) It remains to show $\{C^*_s=\hat{C}^*_s\}$ and $\{C^*_e=\hat{C}^*_e\}$. To accomplish this we need only verify that these choices are feasible at date t=0, since in this case the individual would be willing to move immediately to $\{C^*_s\}$ and $\{C^*_e\}$ at date t=0 and not revise his portfolio at date t=1. Multiplying condition (5) by P_e

$$\Sigma_sP_eP_{s\cdot e}C^*_s+P_eC^*_e=\Sigma_sP_eP_{s\cdot e}\hat{C}^*_s+P_e\hat{C}^*_e$$

and summing over e

$$\Sigma_s(\Sigma_eP_eP_{s\cdot e})C^*_s+\Sigma_eP_eC^*_e=\Sigma_s(\Sigma_eP_eP_{s\cdot e})\hat{C}^*_s+\Sigma_eP_e\hat{C}^*_e$$

By conditions (1) and (4)

$$\Sigma_sP_sC^*_s+\Sigma_eP_eC^*_e=\Sigma_sP_s\hat{C}^*_s+\Sigma_eP_e\hat{C}^*_e=W_0-C_0^*$$

so that

$$\Sigma_s P_s C_s^* + \Sigma_e P_e C_e^* + C_0^* = W_0$$

Consequently, $\{C_s^*\}$ and $\{C_e^*\}$ could have been chosen immediately at date $t=0$.

A similar result has been derived by Jack Hirshleifer and the author in a somewhat less general context. In that paper, it was emphasized that, surprisingly, the nonspeculation condition depends only on the relationship between beliefs and prices, not on resources and tastes.[8] Apparently the variety of consumption claims available in the market at date $t=0$ is sufficiently rich that desired exchange generated solely by differences among individuals in resources and tastes is satisfied by exchange permitted at date $t=0$. In this paper, another implication will be emphasized. Rewriting the nonspeculation condition,

$$\frac{P_{s\cdot e}/P_s}{\pi_{s\cdot e}/\pi_s} = \frac{\pi_e}{P_e}$$

The right-hand side is a number fully determined before state e becomes known; it is a constant of proportionality between revised prices and beliefs, predetermined at date $t=0$, which must hold for no revision to be optimal. Imagine the following vignette:

At date $t=0$, prices $P_e=1/2$ and $P_s=1/20$ are set in the market. From the point of view of one individual $\pi_e=1/2$ and $\pi_s=1/10$. Consequently, $\pi_e/P_e=1$. On the basis of new information received at date $t=1$ under the occurrence of state e, the individual doubles the probability he now attaches to state s ($\pi_{s\cdot e}=1/5$). Before consulting the market, the individual decides to purchase more contingent claims to consumption under state s if its price has not risen sufficiently. However, new information (but not necessarily the same information) may also have become available to other participants in the market, causing them also to revise their beliefs. As a result, a new equilibrium price will be set on the contingent claims. In fact, if the price exactly doubles ($P_{s\cdot e}=1/10$), then the individual will stand pat and not alter his holdings. Why? Because from his point of view, the new information he has received is already fully reflected in the revised price $P_{s\cdot e}$.

If all individuals have homogeneous beliefs, do they all regard the new information they receive at date $t=1$ as fully reflected in revised prices? If the above definition of fully reflected is meaningful, the answer should be positive. Fortunately,

COROLLARY: *If all individuals have homogeneous beliefs, then all individuals have nonspeculative beliefs.*

PROOF:

It follows trivially from the previous theorem that $C_{s\cdot e}^* > (<) \hat{C}_s^*$ if and only if $\pi_s/P_s < (>) \pi_e \pi_{s\cdot e}/P_e P_{s\cdot e}$ (all s and e). If one individual revises his portfolio so that $C_{s\cdot e}^* > \hat{C}_s^*$, by closure, since $\Sigma_i C_{s\cdot e}^i = \Sigma_i \hat{C}_s^i$, there must be some other individual for whom $C_{s\cdot e}^* < \hat{C}_s^*$; but in this case, by the above, the beliefs of both cannot be the same. Hence, the theorem is proved by contradiction.

Although homogeneous beliefs are sufficient for all individuals to see their new information fully reflected in the revised prices, they are not necessary. From the nonspeculation condition, a necessary as well as sufficient condition is that all individuals must have nonspeculative beliefs, that is, agree on the *ratio* $\pi_e \pi_{s\cdot e}/\pi_s$, but with possible disagreement over its components.

For *all* information, not merely new information, to be fully reflected in prices, a stronger condition than nonspeculative beliefs is required. We will say an individual perceives all the information he has and will receive as fully reflected in security prices if and only if he has "consensus beliefs." By definition, *consensus beliefs are those beliefs which, if held by all individuals in an otherwise similar economy, would generate the same equilibrium prices as in the actual heterogeneous economy.* In other words, consensus beliefs are the homogeneous beliefs that would sustain the same equilibrium prices. To motivate this definition, imagine the following vignette:

---

[8] However, while the decision to revise does not depend on resources and tastes, the *extent* of revision does.

> Jack, curious to learn if all his information is fully reflected in prices, visits the Great Economist in the Sky. Knowing the resources and tastes of every participant in the economy, as well as the actual equilibrium prices, the Great Economist calculates whether or not those same prices would rule if every individual had the same beliefs as Jack. Discovering that different equilibrium prices would emerge, the Great Economist concludes that either Jack has valuable information (i.e., it would alter prices if publicly announced) not fully reflected in prices or that other individuals have information affecting prices not known to Jack.

Unfortunately, unlike nonspeculative beliefs, consensus beliefs admit of no simple and general mathematical formulation; they will depend directly on the resources, tastes, and beliefs of all individuals in the economy.[9]

If this interpretation of consensus beliefs is to be meaningful, then surely an individual with consensus beliefs must see his *new* information at date t=1 as fully reflected in revised prices. This is indeed the case since

COROLLARY: *If an individual has consensus beliefs, then he has nonspeculative beliefs.*

PROOF:

By definition of consensus beliefs, an individual who has them can be viewed as a member of an otherwise similar economy, except in which these beliefs are held by everyone. Moreover, by the previous corollary, everyone in a homogeneous economy has nonspeculative beliefs. But since both economies have the same equilibrium prices, nonspeculative beliefs are the same in both economies. Therefore, the individual with consensus beliefs must also

---

[9] This raises the prior question of conditions for the existence of consensus beliefs. In the usual existence problem we input heterogeneous probabilities and solve for nonnegative homogeneous prices. In this case, we input homogeneous prices (which hold in the heterogeneous economy) and solve for homogeneous probabilities (i.e., nonnegative real numbers which sum to one). Unlike prices, these homogeneous probabilities may not exist. While the derivation of necessary and sufficient conditions for their existence is not contemplated here, some sufficient conditions are given as an illustration. In particular, suppose all individuals maximize weighted logarithmic utility functions with the same constant rate of patience; that is, $U_t^i(\cdot)=\rho^t ln(\cdot)$ where $\rho>0$. In this case, it is easily shown that at the optimum $C_0^i=W_0^i(1+\rho+\rho^2)^{-1}$. Substituting this into conditions (2)

$$\pi_e^i \rho C_0^i = C_e^i P_e$$
$$\rho(1+\rho+\rho^2)^{-1}\pi_e^i W_0^i = C_e^i P_e$$

and summing over i and by closure,

$$\rho(1+\rho+\rho^2)^{-1}\Sigma_i \pi_e^i W_0^i = P_e \Sigma_i C_e^i = P_e \Sigma_i \bar{C}_e^i$$

If π_e^M were the consensus belief for state e, then this condition in the homogeneous economy would simplify to

$$\rho(1+\rho+\rho^2)^{-1}\pi_e^M \Sigma_i W_0^i = P_e \Sigma_i \bar{C}_e^i$$

Comparing these two results, the consensus belief

$$\pi_e^M = \Sigma_i \frac{W_0^i}{\Sigma_i W_0^i}\pi_e^i \qquad \text{(all } e\text{)}$$

which is indeed a nonnegative real number such that $\sum_e \pi_e^M=1$. Since similar results hold for conditions (3), and (1), (4), and (5) are trivially satisfied, consensus beliefs exist. In the logarithmic case, they are simply weighted *arithmetic* averages of actual beliefs. See Robert Wilson for illustrations of consensus beliefs if all individuals have exponential utility, in which case they are weighted *geometric* averages of actual beliefs.

have nonspeculative beliefs in the actual heterogeneous economy. (It should be similarly clear that the converse of this corollary is false.)

In sum, we have a triplet of increasingly stronger conditions that capture what is meant by prices fully reflect information: 1) nonspeculative beliefs; 2) consensus beliefs; 3) homogeneous beliefs.

Condition 3) implies 2) and 2) implies 1), but the converses of these implications are false. Condition 3) is sufficient and if consensus beliefs are unique, necessary for *all* individuals to perceive *all* information fully reflected in prices. Condition 2) is necessary and sufficient for *an* individual to perceive *all* his information fully reflected in prices. Condition 1), if true for all individuals, is necessary and sufficient for *all* individuals to perceive *new* information fully reflected in prices; and 1) is necessary and sufficient for *an* individual to perceive his *new* information fully reflected in prices.

Of course, all effects of uncertainty (without assuming it away) on the allocation of resources can be completely exorcised from the economy if individuals have "clairvoyant homogeneous beliefs." By definition, *clairvoyant homogeneous beliefs are beliefs that generate the same allocation of resources as would have emerged in otherwise similar but certain economies, one for each possible sequence of states.* That is, the realized allocation in the uncertain economy is the same allocation that would have emerged in an otherwise similar economy, but where the true sequence of states was known by all individuals in advance of any trading. An economy composed of identical individuals is a trivial instance of clairvoyance, since irrespective of beliefs there is no trading.

III. Rates of Return

The past decade has witnessed the development and testing of theoretical models of the contemporaneous structure of security rates of return under uncertainty. These models, as typified by variants of the popular mean-variance model, describe the relationship among rates of return of *different securities* at the *same date.* Concurrently, there has been considerable empirical research on the intertemporal structure of rates of return of the *same security* at *different dates.* This research has been directed at verification of a variety of statistical hypotheses such as the existence of an unbiased term structure and characteristics of a random walk. However, excepting the incipient work of a very few authors—Jerry Green and John Long—no one has attempted to provide a theoretical basis for these hypotheses in an equilibrium context.

The nonspeculation condition implies restrictions on the intertemporal structure of security rates of return as viewed by individuals who perceive the securities market as informationally efficient. To see this, translating prices into rates of return, let

$$1 + r_{1s\cdot e} \equiv P_{s\cdot e}/P_s$$

$$1 + r_{2s\cdot e} \equiv P_{s\cdot e}^{-1}$$

$$1 + r_{1F} \equiv (\Sigma_e P_e)^{-1}$$

$$1 + r_{2Fe} \equiv (\Sigma_s P_{s\cdot e})^{-1}$$

$$(1 + R_{2F})^2 \equiv (\Sigma_s P_s)^{-1}$$

Using these definitions, we state the following theorem:

THEOREM (Nonspeculative Intertemporal Structure): *In terms of nonspeculative beliefs,*

$$E[(1 + r_{1s})(1 + r_{2s})] = (1 + r_{1F})E[1 + r_{2s}]$$

where the expectations are assessed with respect to beliefs held at date $t = 0$.

PROOF:

By the nonspeculation condition,

$$\frac{\pi_s}{P_s} P_e = \pi_e \frac{\pi_{s\cdot e}}{P_{s\cdot e}}$$

Summing over states e and rearranging

$$\pi_s P_s^{-1} = (\Sigma_e P_e)^{-1}(\Sigma_e \pi_e \pi_{s\cdot e} P_{s\cdot e}^{-1})$$

Since $\pi_s = \Sigma_e \pi_e \pi_{s\cdot e}$,

$$\Sigma_e \pi_e \pi_{s\cdot e}\left(\frac{P_{s\cdot e}}{P_s} P_{s\cdot e}^{-1}\right) = (\Sigma_e P_e)^{-1}(\Sigma_e \pi_e \pi_{s\cdot e} P_{s\cdot e}^{-1})$$

The conclusion follows immediately upon substituting rate of return notation.

In words, the intertemporal structure of rates of return implied by nonspeculative beliefs (and hence by consensus or homogeneous beliefs) must have the following characterization: *the expected one plus compound rate of return on any security discounted by one plus the first-period risk-free rate equals its expected one plus second-period rate of return, where the expectations are assessed with respect to beliefs at the beginning of the first period.*[10]

If we assume this minimal characterization of the intertemporal structure, can we confirm or reject any popular statistical hypotheses? Some trial hypotheses are listed below:

(A) **No Risk Adjustment**

$$1 + r_{1F} = E[1 + r_{1s}]$$

(B) **Uncorrelated Rates**

$$E[(1 + r_{1s})(1 + r_{2s})] = E[1 + r_{1s}]E[1 + r_{2s}]$$

(C) **Uncorrelated Inverse Rates**

$$E[(1 + r_{1s})^{-1}(1 + r_{2s})^{-1}] = E[(1 + r_{1s})^{-1}]E[(1 + r_{2s})^{-1}]$$

(D) **Unbiased Inverse Term Structure**

$$(1 + R_{2F})^{-2} = (1 + r_{1F})^{-1}E[(1 + r_{2F})^{-1}]$$

It follows immediately from the previous theorem that in terms of nonspeculative beliefs, (A) and (B) are equivalent. That is, *the expected first-period rate of return on any security equals the first-period risk-free rate if and only if its intertemporal sequence of one plus rates of return are serially uncorrelated.* Consequently, we cannot simultaneously believe in a risky economy composed of risk-averse individuals and observe *ex ante* serially uncorrelated (and, hence, observe serially independent) one plus rates of return. However, although hypotheses (A) and (B) may be rejected, neither hypothesis (C) nor (D) can be confirmed or rejected in the presence of nonspeculative beliefs. Serially uncorrelated rates of return, as, for example, surveyed in Fama, have played a significant role in the guise of the random walk hypothesis in statistical tests of information efficiency. However, the theoretical analysis provided here indicates that at least some forms of the random walk hypothesis are inconsistent with information efficiency. To obtain a more specific characterization of the intertemporal structure of security rates of return, it is necessary to impose stronger assumptions than mere nonspeculation.

IV. Concepts of Information Efficiency

It is instructive to contrast the concept of information efficiency developed in this paper with other discussions in the literature. To paraphrase Fama, efficiency is defined as follows:

> Using information set ϕ, forecast the probability distribution of the prices of securities that will be realized at date t=1. Input this data into a model of market equilibrium determining expected rates of return. From these and the expected date t=1 prices, the date t=0 prices, *given* information set ϕ *and*

---

[10] Caution: This characterization will in general hold only for rates of return of each contingent claim, not for rates of return on *portfolios* of contingent claims.

> the market equilibrium model, can be computed. Compare these computed prices with the prices actually observed in the marketplace. If they are the same, then actual security prices are said to "fully reflect" ϕ. [p. 384]

Although less precise, the implied beliefs of this definition appear similar to consensus beliefs. However, Fama's definition and subsequent use of it suffer from several deficiencies. First, his definition is unnecessarily couched in terms of expected returns. Second, suppose ϕ is fully reflected by his definition and an individual knows only some proper subset of ϕ. He may then assess different date t=0 prices and, although he will be unable to make superior returns relative to ϕ, his mistaken estimate of date t=0 prices will lead him to perceive some securities as over- and undervalued. As a consequence, he may neglect to diversify efficiently. By Fama's definition, an efficient market may coexist with this nonoptimal behavior. Third, unable to draw nontrivial, testable implications from his definition, Fama is forced to supplement it with highly specialized and unrealistic models of equilibrium.

By contrast, the definition developed in this paper arises from within an explicit but quite general equilibrium context. Despite its generality, precise characteristics are derived describing individuals who view the market as efficient (i.e., nonspeculation condition). An important distinction is made between new information and all information. The definition is also linked with the volume of trading. For example, if every individual viewed the market as efficient, there would be no trading at date t=1. In a more general model, with the continuous emergence of new securities and new participants as well as barriers to marketability, the model holds out the promise of distinguishing trading volume for these reasons from trading for speculative purposes. The volume of speculative trading could then be used as a barometer of information efficiency.[11]

Samuelson attempts to justify martingale characterizations of efficiency. He proves the following exceedingly simple theorem:[12]

THEOREM (Fair Game): *If* $P_s=\pi_s$ *and* $P_{s\cdot e}=\pi_{s\cdot e}$ *(all e), then* $P_s=\Sigma_e\pi_e P_{s\cdot e}$.

PROOF:

By the laws of probability, $\pi_s=\Sigma_e\pi_e\pi_{s\cdot e}$ and by assumption $P_s=\pi_s$ and $P_{s\cdot e}=\pi_{s\cdot e}$ (all *e*).

The term P_s, it will be recalled, is the date t=0 price of one unit of consumption received at date t=2 if and only if state *s* occurs. Since the date t=2 price of one unit of consumption received at date t=2 is just 1, π_s may be interpreted as the expected date t=2 price of one unit of consumption received at date t=2 if and only if state *s* occurs, where the expectation is assessed with respect to beliefs held at date t=0. Similarly, given that state *e* has occurred, $P_{s\cdot e}$ is the date t=1 price of one unit of consumption received at date t=2 if and only if state *s* occurs. Given that state *e* has occurred, $\pi_{s\cdot e}$ may be interpreted as the expected date t=2 price of one unit of consumption received at date t=2 if and only if state *s* occurs, where the expectation is assessed with respect to beliefs held at date t=1. Loosely speaking, then, the fair game theorem says: *If the price of a security at every date always equals its expected price at some terminal date, then its price at every date equals its expected price at any future date prior to the*

---

[11] David Downes and Thomas Dyckman, pp. 314–16, offer a less specific but similar suggestion.

[12] Samuelson actually proves a more general version that allows for continuous probability distributions, any finite number of dates, and incomplete markets. However, nothing of significance is lost in the more straightforward theorem given here.

terminal date. This stochastic process of price behavior is called a martingale.

The theorem is obvious. Since its assumption ignores time preference and risk aversion,[13] it is not surprising to learn that the date t=0 price of a consumption claim equals its expected date t=1 price. As it stands, the theorem is a trivial characterization of efficiency, or in Samuelson's words, "properly anticipated prices." Possibly realizing this, Samuelson generalizes the result to allow for positive risk-free interest and risk aversion:

COROLLARY: *If* $P_s=\lambda_{1s}^{-1}\lambda_{2s}^{-1}\pi_s$ *and* $P_{s\cdot e}=\lambda_{2s}^{-1}\pi_{s\cdot e}$ *(all e), then* $P_s=\lambda_{1s}^{-1}\Sigma_e\pi_e P_{s\cdot e}$.

λ_{1s} and λ_{2s} are time and risk-adjusted discount rates that may be different for different consumption claims (i.e., depend on state *s*), but that are ***independent of state e***. Indeed, in this lies the content of the theorem, since if λ_{2s}^{-1} in the second equality were replaced by $\lambda_{2s\cdot e}^{-1}$, the assumption of the corollary would be vacuously true. Examination of the proof of the fair game theorem provides an immediate proof of this generalization.

What is the actual content of these results and how are they related to the definition of efficiency provided in this paper? In view of the restrictive equilibrium implications of the fair game assumptions, we will turn more hopefully to Samuelson's generalization. However, in the context of nonspeculative beliefs (and, hence, consensus or homogeneous beliefs), a significant inconsistency emerges. By the assumption of his corollary, $\pi_s/P_s=\lambda_{1s}\lambda_{2s}$ and $\pi_{s\cdot e}/P_{s\cdot e}=\lambda_{2s}$ (all *e*). Substituting these relationships into the nonspeculation condition $\pi_s/P_s=(\pi_e\pi_{s\cdot e})/(P_e P_{s\cdot e})$ (all *e*), we derive $\pi_e/P_e=\lambda_{1s}$ (all *e*). This not only means π_e/P_e must be independent of *e*, but also that λ_{1s} must be independent of *s*. Indeed, by summing over *e*, $\lambda_{1s}=1/(\Sigma_e P_e)$. Interpreting this verbally, all claims to consumption at date t=2 have the same risk-adjusted discount rate between dates t=0 and t=1 equal to one plus the risk-free rate. In effect, despite Samuelson's attempt to generalize to risk-adjusted discount rates, an individual with nonspeculative (or consensus, or homogeneous) beliefs will use the risk-free rate, not adjusting for risk.

As Samuelson asks at the end of his paper:

> I have not here discussed where the basic probability distributions are supposed to come from. In whose minds are they ex ante? . . . Are they supposed to belong to the market as a whole? And what does that mean? Are they supposed to belong to the "representative individual," and who is he? . . . This paper has not attempted to pronounce on these interesting questions. [pp. 48–49]

While the present paper has not unveiled this mystery man, if securities are not all to be discounted at the risk-free rate, he cannot have homogeneous, consensus, or even nonspeculative beliefs. In short, he may not be an economically interesting individual.

Another tempting definition is to associate with information efficiency the holding of (possibly levered) value-weighted balanced portfolios. However, we know from David Cass and Stiglitz that even in a static homogeneous economy, individuals will not generally desire to hold (possibly levered) miniatures of the market portfolio. Yet, by any meaningful definition, homogeneous belief is surely a sufficient condition for information efficiency.

"Intrinsic value" is another recurrent phrase associated with efficiency. Clearly, nothing has intrinsic value apart from the resources, tastes, and beliefs of all individuals in the economy. However, taking a more charitable view toward intrinsic value, perhaps it refers to valuations based on the pool of all available information. In other words, *intrinsic values are those equilibrium prices set as if each individual publicized all his information.* In this "as if" economy, all individuals would have homogeneous beliefs. If *these as if* homogeneous beliefs happen to coincide with the actual consensus beliefs, only then, by this definition, will the actual market be efficient. Although this definition is perhaps closest in spirit to the implicit definitions used in most of the empirical literature, it still suffers from the coexistence of nonoptimal speculative trading. That is, just because equilibrium prices are set as if each individual publicized all his information does not mean that any individual bases his choices on the pooled set of available information.

---

[13] Alternatively, it ignores time preference and presumes a riskless economy in which risky securities may exist as long as, through offsetting effects, aggregate consumption is certain.

V. Limitations

The chief difficulty with the concept of information efficiency developed in this paper has been the absence of transactions costs. Without these costs, in a heterogeneous economy, complete markets are a natural evolution of the structure of exchange opportunities. At a deeper level, the choice, cost, and the communication of information which creates transactions cost has not received even implicit attention.

It is useful to distinguish between two types of information about security prices —information about the *supply* of aggregate consumption and information about the *demand* for it. The dynamic optimization problem envisaged here requires a considerable degree of forethought: an individual must not only choose (C_0, $\{\hat{C}_e\}$, $\{\hat{C}_{s \cdot e}\}$), but also decide in advance at date t=0 how he will revise these provisional choices at date t=1 (choose (C_e, $\{C_{s \cdot e}\}$) for all states e). In this planning process, at date t=0 the individual utilizes date t=1 equilibrium state conditional prices even before the date t=1 market convenes! In effect, we have assumed that given the resolution of uncertainty surrounding the aggregate supply of consumption, demand conditions and hence prices are known with certainty.[14] In a more general setting, even if supply uncertainty were resolved, lacking information about the resources, beliefs, and tastes of other individuals, an individual could not forecast demand and hence prices with certainty. Unfortunately, for issues examined in this paper, absence of demand uncertainty is particularly objectionable since it cavalierly ignores the Keynesian or psychological view of the securities market in which speculative bubbles are a possibility.[15]

While the reduction of demand uncertainty motivates transactions costs, supply uncertainty also motivates the acquisition of information. The model developed in this paper contained no explicit analysis of its choice—beliefs were taken as data. Moreover, although different individuals had different beliefs, this information was implicitly assumed to be costless. With explicit consideration of productive choice and delegation of authority to managers of firms, supply uncertainty would become more complex. Individuals would not only be uncertain about the outcomes of "natural processes," but also about the behavior of managers. In turn, this additional cause of uncertainty would derive from two sources: uncertainty about their reported production choices and uncertainty about discrepancies between their actual and reported choices.

---

[14] Actually, this statement only holds for results, such as the identification of consensus beliefs, which require a market equilibrium setting. The identification of nonspeculative beliefs requires only a model of individual behavior in which uncertainty about future prices can derive from both supply and demand conditions.

[15] This distinction between supply and demand uncertainty would seem to be at the root of the separation of so-called "fundamental" and "technical" approaches to security analysis.

A theoretical exploration of the meaning of the phrase "security prices fully reflect information," in the presence of the choice, cost, and communication of information potentially capable of reducing demand and supply uncertainty, must await another opportunity.

REFERENCES

K. J. Arrow, "The Role of Securities in the Optimal Allocation of Risk Bearing," *Rev. Econ. Stud.*, Apr. 1964, *31*, 91–96.

D. Cass and J. E. Stiglitz, "The Structure of Investor Preferences and Asset Returns, and Separability in Portfolio Allocation," *J. Econ. Theory*, June 1970, *2*, 122–60.

G. Debreu, *Theory of Value*, New York 1959.

D. Downes and T. R. Dyckman, "A Critical Look at the Efficient Market Empirical Research Literature as It Relates to Accounting Information," *Accounting Rev.*, Apr. 1973, *48*, 300–17.

J. Dreze, "Market Allocation Under Uncertainty," *European Econ. Rev.*, Winter 1970, *2*, 133–65.

E. F. Fama, "Efficient Capital Markets: A Review of Theory and Empirical Work," *J. Finance*, May 1970, *25*, 383–417.

J. Green, "A Simple General Equilibrium Model of the Term Structure of Interest Rates," disc. pap. no. 183, Harvard Inst. of Econ. Res., Mar. 1971.

N. H. Hakansson, "Optimal Entrepreneurial Decisions in a Completely Stochastic Environment," *Management Sci.*, Mar. 1971, *17*, 427–49.

J. Hirshleifer and M. Rubinstein, "Speculation and Information in Securities Markets," *Proc. Inst. Management Sci. 1973*, Vol. 2, Tel-Aviv 1975, 799–802.

J. Long, "Consumption-Investment Decisions and Equilibrium in the Securities Market," in M. Jensen, ed., *Studies in the Theory of Capital Markets*, New York 1972.

M. Rubinstein, "Jan Mossin's *Theory of Financial Markets*," *Bell J. Econ.*, Autumn 1973, *4*, 693–99.

———, "An Aggregation Theorem for Securities Markets," *J. Financial Econ.*, Sept. 1974, *1*, 225–44.

P. A. Samuelson, "Proof That Properly Anticipated Prices Fluctuate Randomly," *Ind. Manage. Rev.*, Spring 1965, *6*, 41–50.

J. E. Stiglitz, "A Consumption-Oriented Theory of the Demand for Financial Assets and the Term Structure of Interest Rates," *Rev. Econ. Stud.*, July 1970, *37*, 321–51.

R. Wilson, "The Theory of Syndicates," *Econometrica*, Jan. 1968, *36*, 119–32.

The Consumption Function in a Developing Economy and The Italian Experience

By F. MODIGLIANI AND E. TARANTELLI*

The many models that have been proposed since J. M. Keynes' *General Theory* to explain aggregate consumption and saving have been largely developed with reference to mature, western-type, capitalistic economies. This paper examines, in the first place, how well the various models can account for the behavior of saving in a country undergoing rapid development such as Italy in the postwar period. It next reviews a number of "stylized facts" characterizing the process of development, which, in turn, lead to some needed extensions of existing models.

Our main conclusions can be summarized as follows: 1) All major models, starting from the standard textbook consumption function, provide a reasonably good approximation to the behavior of aggregate consumption in the two decades terminating in the early 1970's. At the same time, the successive vintages of the consumption function do yield some improvements. 2) On the whole, the model that appears to provide the most fruitful framework for a consistent understanding of saving behavior, both at the aggregate and the cross-sectional level, is the life cycle hypothesis of consumption (*LCH*) in its more general form, according to which consumption depends on expected income from labor, wealth, and "permanent" income from wealth. 3) However, even this model needs to be modified along Ricardian-Marxian lines to take account of the consumption behavior of the reserve army of the structurally unemployed, which gradually dwindles in the process of development. 4) Our other major result is to provide overwhelming evidence against the standard Kaldorian two-casts parable of hard-pressed workers consuming all and plush capitalists saving almost all, which so far had been held to provide the most satisfactory explanation of the behavior of aggregate saving in Italy. 5) The crucial piece of evidence supporting the proposition stated in (4) is that the consumption coefficient of "expected" profits is not only larger than that of expected labor income, but is in fact larger than one. This result is not merely inconsistent with the Kaldorian hypothesis but altogether uninterpretable in that framework, whereas it is shown to be fully consistent with the implications of the life cycle model. In this model every household is typically both a worker and a recipient of property income during at least some phase of the life cycle and may be expected to respond differently to these two differing sources of income. 6) It is shown that the coefficient measuring the response of consumption to property income cannot be interpreted as a marginal propensity in the customary sense of the word and reflects, at least in part, the degree of substitution between current and future consumption in the utility theoretic sense.

---

* Massachusetts Institute of Technology. Some of the equations proposed here are part of the Econometric Model of the Bank of Italy, which partly supported the present work, and whose aid is gratefully acknowledged. We also wish to thank the National Bureau of Economic Research for use of the TROLL System for Applied Econometric Research, and, finally, to express our appreciation to many colleagues of the Research Division of the Bank of Italy for useful discussion and advice. In particular Bruno Sitzia contributed importantly to the development of the tests discussed in Section IV. The main issues examined in this paper are dealt with in greater detail in our forthcoming book.

I. The "Old Theories," the "Permanent Income" Hypothesis, and the Kaldorian Model

The first conclusion listed in the introduction can be readily established by an inspection of Table 1 which shows the result obtained by fitting to Italian aggregate data the standard Keynesian model (equation 1), the Duesenberry-Modigliani (D-M) 1949 model in its original form (equation 2), and the Tillman Brown-Tom Davis (B-D) version (equation 3). Strictly speaking, the explanatory variable called for by the D-M model is, besides current income, the highest previous income. But, in the case of Italy, as for any continuously growing economy, the highest previous income reduces to lagged income. Similarly, for the B-D model the highest previous consumption coincides with lagged consumption.

For every model, two alternative estimates of the parameters are provided, first using the linear form (equations labeled L),

TABLE 1—ESTIMATES OF PARAMETERS OF CONSUMPTION FUNCTION: ITALY, 1952–70

Consumption Function		Coefficients of						
		Constant	Y	Y_{-1}	C_{-1}	$P(Y/YN)$	$S.E.$	$D.W.$
Elementary Keynes	1L	1355.4	.78				322.3	1.41
		(246.8)	(.011)					
	1R	1687.7	.76				.0154	.98
		(224.9)	(.012)					
Duesenberry Modigliani (D-M)	2L	1254.3	.23	.58			250.3	1.23
		(193.8)	(.16)	(.17)				
	2R	1571.5	.33	.46			.0134	.95
		(200.5)	(.17)	(.18)				
Brown-Davis Friedman (B-D)	3L	499.1	.40		.52		260.9	1.66
		(337.1)	(.12)		(.17)			
	3R	648.5	.37		.55		.0122	1.32
		(356.9)	(.12)		(.16)			
Kaldor (KH)	4L	2927.2	1.02			−.70	245.1	1.17
		(468.5)	(.068)			(.19)		
	4R	3448.9	1.06			−.83	.0106	1.07
		(421.5)	(.067)			(.19)		

Note: The figure in parentheses underneath each coefficient is the standard error of the coefficient. $S.E.$ and $D.W.$ denote, respectively, the standard error of the regression and the Durbin-Watson coefficient. C, Y, and $P(Y/YN)$ denote, respectively, total private domestic consumption, total disposable income, and "independent labor income" (national income less wages and salaries) multiplied by the ratio of disposable to net national income. All variables were measured in billions of lire at 1963 prices. (Over the period of analysis, 1 billion lire was roughly equivalent to 1.6 million dollars at the official exchange rate.) The average value for C in the period under consideration was 18,340 billions of lire. Dividing $S.E.$ by this value one can calculate the percentage standard error of the linear equations. For the equations expressed in ratio form (R), the dependent variable is C/Y, and $S.E.$ gives the root mean square error as a percentage of income. Disposable income was defined as national income plus government transfers minus personal and corporate income taxes. Note that this definition differs from the standard one because it includes undistributed corporate profits.

and second dividing both sides of the equation by income (equations labeled R). This alternative estimate is provided because, over the period of observation, both consumption and income have grown very substantially (roughly by a factor of three) and, therefore, the error term may be more nearly homoscedastic in the ratio than in the linear form. Note also that the models tested would generally call for the use of per capita rather than aggregate real data. However, since population has changed very little over the period of observation, we find it convenient to use aggregate data, expressed in real terms using as a deflator the implicit price index of private domestic consumption. It is apparent that even the standard Keynesian function fits well, as its standard error is well below 2 percent of consumption (see the footnote to Table 1). The next two models yield appreciably smaller standard errors. Note also that in terms of fit, the B-D model is an improvement over the D-M model only in its ratio version. However, it does yield a much smaller and barely significant constant term which accords better with the specification of the hypothesis and the nowadays widely held view that the saving ratio should, in the long run, be independent of the level of income (though not of its rate of growth). Note that in both the D-M and B-D equations the coefficient of the lagged variable is distinctly larger than that of the current variable, implying a long lag in the response of consumption which will play an important role in what follows.

The three models discussed so far belong to the empirically oriented tradition of the early post-Keynesian period, that is, they belong to what has since come to be known, in Michael Farrell's words, as the "old theories of the consumption function." However, equation 3 can also be interpreted as providing a test of one version of the "new theories," namely, Milton Friedman's Permanent Income Hypothesis. As is well known, Friedman suggested that permanent income could be approximated by an exponentially weighted average of past incomes, a formulation which, making use of the L. M. Koyck transformation, leads to an autoregressive equation of the form of equation 3.

Equation 4 provides a test of Nicholas Kaldor's hypothesis (hereafter KH) that the propensity to consume of the recipients of property income, the "capitalists," is significantly lower than that of the recipients of labor income, the "workers." Denoting by a_p the propensity to consume out of net of tax property income P^*, and by a_w the corresponding propensity for wage income W^*, KH can be stated as

$$(4) \qquad C = a_w W^* + a_p P^*, \quad a_p < a_w$$

Recalling that $W^* + P^* = Y$, (4) can be rewritten in ratio form as

$$(4') \qquad C/Y = a_w + (a_p - a_w)P^*/Y$$

which shows that the saving ratio $S/Y = 1 - C/Y$ does not depend on income as such, but instead on income distribution, P^*/Y.

The national income accounts do not provide explicit measures of W^* and P^*, i.e., labor and property income net of taxes, but provide only estimates of *dependent* labor income (wages and salaries) and of other incomes (with various breakdowns) gross of personal taxes. These variables have, therefore, been approximated by relying on the assumption that the ratio between P^* and W^* and the corresponding measures gross of taxes, P and W, are approximately equal, i.e., that the incidence of taxes is roughly the same on the two sources of income. This assumption, though risky, finds some justification in the tax structure of the country taking into account the phenomenon of widespread tax evasion, especially on property income. Thus,

$$\frac{P^*}{P} \simeq \frac{W^*}{W} \simeq \frac{Y}{YN} \tag{5}$$

where YN is "real" (net) national income. From (4) and (5) and since $W+P=YN$, one derives

$$C = (a_p - a_w)P\frac{Y}{YN} + a_wY \tag{6}$$

Following studies made by A. Niccoli and S. Vinci, for the Bank of Italy in 1970, testing KH for the Italian economy, a constant was added to (6) even though it presumably should be close to zero. It is apparent from equation 4 that both the scaled and unscaled equations fit the data somewhat better than the previous models, though in both equations the Durbin-Watson statistic still indicates the existence of substantial positive serial correlation.

These results, and similar ones obtained by others, have often been interpreted as providing strong support for the Kaldorian model. They imply, in fact, a marginal propensity to consume out of labor income of 1.02 to 1.06, and corresponding propensities to consume out of property income of only .32 to .23. Yet these estimates show a number of grave inconsistencies and dubious implications. First, the estimated marginal propensity to consume out of wage income appears unreasonably high. Second, the estimated marginal propensity to consume for nonlabor income as low as .23—i.e., four to five times lower than that out of wages—is rather dubious. It is, in fact, glaringly inconsistent with the results of budget studies.

One of the most complete and comprehensive series of these studies for the Italian economy, conducted by A. Ulizzi for the Bank of Italy, classifies families into two groups, dependent workers and the self-employed, which correspond to those underlying the National Income classification of dependent and independent workers. The behavior of savings within each class is very much in line with what one would expect on the basis of the permanent income or life cycle hypothesis of savings. Specifically, for the dependent families whose income may be presumed to be relatively stable over time, the proportion of income saved is roughly independent of income and around 7 to 8 percent (see Ulizzi, p. 424, Table 3.4 and Graph 2), at least for incomes above 1 million lire. (Even for income recipients below 1 million, which undoubtedly contain a sizable proportion of the transiently poor, saving is positive and around 3 percent.) For the group as a whole, average saving amounts to 5.8 percent, which is clearly inconsistent with the results given above implying a *negative* average propensity to save of 2 to 6 percent. On the other hand, for the self-employed, including farmers and other entrepreneurs whose income is notoriously subject to wide short-run fluctuations, the saving ratio tends to rise markedly with income, being negative for incomes up to 2 million lire and increasingly positive for the remaining two groups. This is precisely what one would expect for a group whose income is uncertain and fluctuating, if consumption depends on "permanent income" (Friedman), or "life resources" (Modigliani-Ando, 1960). For such groups, a better estimate of the average and marginal propensity to save is provided by the sample average which turns out to be 8.4 percent.[1] These results indicate a roughly similar average propensity to consume for the two groups, and contradict those suggested by the Kaldorian equation (4).

---

[1] These figures were kindly provided by Ulizzi and are consistent with the results shown in Table 3.4 of his work. In his definition, the self-employed include workers in agriculture and other sectors, entrepreneurs, and professionals. Dependent labor includes wage earners and employees. Pensioners are excluded from both definitions of income. This, together with other characteristics of the sampling procedure adopted, leads to an underestimate of the average propensity to save which should not, however, appreciably affect the magnitude of the propensities to save relative to each other.

Finally, in equation (4), the constant term is by far the highest of all the estimates presented so far and is statistically highly significant. This result plays havoc with the well-known Kaldorian theory of income distribution. Indeed, adding a constant a to (4), the Kaldorian saving function can be written as: $S=(1-a_w)W^*+(1-a_p)P^*-a=I$, where I denotes investment. Substituting for W^* from the identity $Y=W^*+P^*$ and solving for P^*, the share of profits implied by the Kaldorian theory of income distribution can be written as:

$$\frac{P^*}{Y}=\frac{1}{(a_w-a_p)}\left[\frac{I}{Y}-(1-a_w)+\frac{a}{Y}\right]$$

which, in turn, implies that given a_p and a_w, and for a given investment ratio, the long-run share of profits is a decreasing function of income and not a constant as implied by the theory. (See Kaldor, p. 229.)

It is the contention of this paper that the improbable estimates of the parameters of (6) result from the failure to take into account a number of fundamental characteristics of a developing economy, like Italy's, in the period under consideration and their implications for an understanding of consumption behavior. These characteristics are reviewed below.

II. Some "Stylized Facts"

A

First, it is wrong to identify dependent labor income with Kaldorian labor income and the remaining income with profits, because of the well-known fact that in an undeveloped economy such as Italy's in the early postwar years, the *independent* labor pool contained a substantial portion of self-employed, of marginally employed, and/or of "disguised" unemployed. Indeed, in 1951, the number of supposedly "capitalist" independent workers amounted to 7,700,000 on a total labor force of some 17 million.[2] Furthermore, in the early 1950's, the per capita income of the independent workers was quite close to that of the dependent workers as shown by Figure 1, which traces the behavior of the ratio of per capita income of dependent workers to that of the independent workers. Thus, in 1951, per capita dependent labor income equalled $720 as against only $890 for per capita "profits." (In other words, according to the conventional definition, the capitalists would have shared the misery of the workers!)

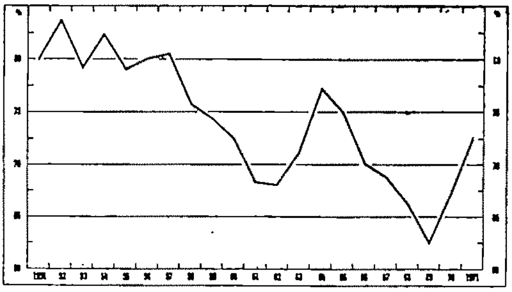


Figure 1. Ratio Between Per Capita Dependent Labor Income and Other Incomes

As also indicated by the budget studies mentioned previously, these small differences in the average per capita earnings of the two groups cannot account for the extreme gap in the two propensities to consume suggested by the "naive" Kaldorian equation. The narrow differences in earnings[3] suggest, in fact, that there has been and still is a whole group, conventionally included in the independent labor category, for which average per capita income and saving behavior are likely to be roughly similar to those of the dependent labor group.[4]

---

[2] We have been told that in some parts of the south, this group of people is jokingly referred to as "gli industriali, perchè si industriano" which can be roughly paraphrased as "the managers, because they manage." For a stimulating and enlightening discussion of the "class composition" of the Italian labor force and its development in time, see P. Sylos-Labini (1972).

[3] In 1971, per capita dependent "labor income" averaged $4000, as against only $5400 for per capita "profits." See Tarantelli (1973a, b).

B

The measured share of wages and salaries as a proportion of net national income has increased remarkably in the period under consideration. As the Annual Report of the Bank of Italy for 1971 has shown,[5] this change does not arise from an increase in the average per capita earnings of this group relative to the income of independent labor: the ratio between the two has, in fact, fallen, as shown by Figure 1. It arises instead from the increase in the share of dependent labor in the total labor force, as one might expect for a developing economy,[6] as a consequence of two major features of development: a) the continuous flow of marginal and low-income independent workers into the dependent labor pool; b) the substantial fall, up to the beginning of the 1960's, in the number of structurally unemployed who were also largely absorbed into the dependent labor pool. The report of the Bank of Italy also suggests that once the data are "purged" of the bias due to the change in the composition of employment, the long-run trend in favor of wages shown by Figure 2 disappears. The "corrected" distribution of income shows modest cyclical fluctuations around a constant, as shown by Figure 3.

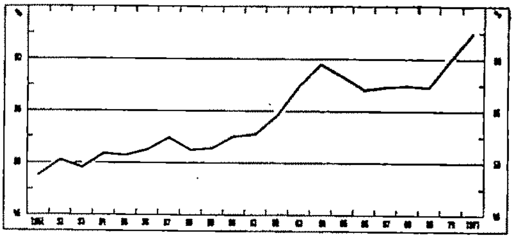


FIGURE 2. SHARE OF WAGES AS A PROPORTION OF NET NATIONAL INCOME

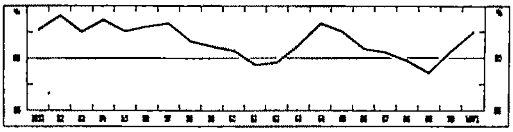


FIGURE 3. WAGE SHARE "CORRECTED" FOR THE VARIATIONS IN THE COMPOSITION OF EMPLOYMENT

C

It seems plausible to assume, à la Karl Marx, that the marginal and average propensity to consume of the structurally unemployed in a developing economy equals one. This group is a receiver of dependent and independent labor income from occasional gainful employment and from low earnings in unrecorded activities and of transfer income from the gainfully employed.

Because of the higher propensity to consume of the structurally unemployed, one would expect that in a developing economy, the gradual absorption of this group into the labor pool would involve a corresponding fall in the average propensity to consume of the system as a whole.[7] This presumption is fully verified by the Italian postwar experience. As Figure 4 shows, the average propensity to consume has substantially fallen from 1951 to the beginning of the 1960's, while following a cyclical pattern thereafter. Figure 4 also shows the remarkable fall in the number of the structurally unemployed in the 1950's. The parallelism of the two series in the graph would seem to support the

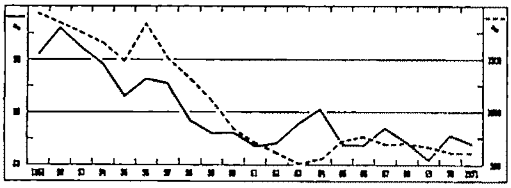
---

[4] Along similar lines, Sylos-Labini (1967) suggested that "dependent workers have a propensity to consume near to unity, that the independent workers receive an income that moves together with that of the dependent ones and have a propensity similarly high" (p. 360).

[5] See Banca d'Italia (1971, p. 147).

[6] By 1971, the number of independent workers had shrunk from 7.7 to 5.7 million, representing 30 percent of the labor force, as against 46 percent in 1915.

[7] For earlier interest in this evidence, see F. Di Fenizio and A. Graziani.



– – – – – – Number of Unemployed (in thousands)
———————— Average Propensity to Consume

FIGURE 4

Ricardian-Marxism hypothesis proposed above.[8]

These "stylized facts" suggest two major modifications of (6). First, following the procedure of the Bank of Italy, total labor income is estimated as the sum of dependent labor income and an ***imputed*** labor income of the ***independent*** workers measuring, in essence, their "opportunity cost" of self-employment. This opportunity cost is assumed to equal, on the average, the average wage income of dependent workers W/LD, where LD stands for the number of dependent workers. Thus, letting LL denote total employment, our measure of the "corrected net of tax labor income" becomes $(W/LD)LL(Y/YN)$.[9] Second, one needs to account for subsistence consumption of the dwindling army of the unemployed, U.[10] Thus, equation (6) must be modified to:

$$C = a_w\hat{W} + a_p\hat{P} + a_uU \tag{7}$$

where U denotes the number of unemployed and

$$\hat{W} = \frac{W}{LD}LL\frac{Y}{YN}; \quad \hat{P} = Y - \hat{W} \tag{7'}$$

Equation (7) can be used directly for linear estimation, and the coefficients a_p and a_w of $\hat{P}$ and $\hat{W}$ provide direct estimates of the Kaldorian marginal propensity to consume (MPC) of the capitalists and the rest of the population, respectively. For ratio estimation, in order to preserve the same dependent variable used in the test of the other models, namely C/Y, it is necessary to perform a transformation by adding and subtracting $a_p\hat{W}$, which yields

$$\frac{C}{Y} = a_p + (a_w - a_p)\frac{\hat{W}}{Y} + a_u\frac{U}{Y} \tag{7''}$$

III. Empirical Test of the Modified Kaldorian Model

Empirical tests of equations (7) and

---

[8] It seems worth noting that the recognition of the above *stylized facts* can readily account for the relatively good mechanical fit of the naive Kaldorian equation 4 and some of its properties. The large constant term in equations 4L and 4R of Table 1 mechanically accounts for the higher than average and falling average propensity to consume up to the beginning of the 1960's. To be sure, this constant would imply, on the other hand, a fall in the average propensity to consume also for the rest of the period under consideration, contrary to what has in fact happened. Equations 4L and 4R account for this pattern by assigning an unreasonably high weight to the share of wages, which as can be seen from Figure 2, increased but little up to 1960 (from .49 to .52), but substantially in the second half of the period (from .52 to .61), thus offsetting the depressing effect of the constant term.

[9] This approach assumes that there is no appreciable number of "pure capitalists." This conjecture is supported by the estimates produced by Sylos-Labini (1972). In any event, we have also computed an alternative estimate of labor income assuming that pure capitalists represented 30 percent of employment, the proportion of self-employed at the close of the sample period. To this end, the average income of dependent labor was imputed to only 70 percent of the labor force. This alternative was used in all the tests reported below, but since in no case did it lead to qualitatively different conclusions, the results (and their implications for shares of the dependent labor force included between these two alternatives) are omitted here, though they can be found in the authors' forthcoming book. The measure used here also coincides with that discussed in Tarantelli (1970). See also Tarantelli (1973a,b).

[10] As shown in the more extensive version of our model cited earlier, the unemployment component should in principle distinguish between structural and cyclical unemployment and account for the variations of the average income per capita of dependent labor. See also the same reference for a discussion of the issues involved and for a detailed derivation of equation (7) from the aggregation of the relevant income subgroups. The distinction between structural and cyclical unemployment is also elaborated by the authors (1973).

(7″) are reported in the equations labeled 1 and 2, Table 2. In equation 1, we omit the unemployment variable, mainly for purposes of comparison with the standard Kaldorian equation reported in Table 1, 4*L* and 4*R*. While the estimated coefficients have the expected sign, all equations give a somewhat poorer fit than the equations of Table 1. The estimated *MPC* out of wages falls to a more reasonable .85, while that out of profits is unaffected. But the constant terms remain highly significant in all cases.[11]

The addition of the variable relating to structural unemployment in equation 2 improves substantially the general characteristics of the equations. The constant becomes of negligible magnitude and statistically insignificant, implying a long-run consumption-income ratio independent of income for given distributive shares. As a result, if the constant term is suppressed as is shown in equations 3*L* and 3*R*, the coefficients of the variables are essentially unaffected and the standard error adjusted for degrees of freedom is actually reduced. Similar results were found to hold also for the other versions of the model reported in Table 2. In view of this, and since a priori considerations suggested that the constant term should in fact be zero, further results are reported only for equations estimated constraining the constant term to be zero.

The order of magnitude of the point estimates of the coefficient of U implies an average "subsistence minimum" slightly above half a million lire ($800) which is somewhat high, but not altogether beyond reason. However, the estimated marginal propensity of "capitalists" remains rather small, .42, still only half as large as a_w. Thus the results of row 3 would seem to provide strong support for Kaldor's model.

But these equations still suffer from one serious shortcoming, namely that they make no allowance for the possibility of a gradual response of consumption to income, especially property income, as suggested by a number of earlier studies on this topic and by previous considerations. This could, in part at least, explain the relatively low *t*-ratio of property income, and the autocorrelations of residuals in equation 3 of Table 2.

An operationally simple approach to allow for this possibility is to amend (7) as follows

$$(8)\ C = a_w\hat{W} + a_p\hat{P} + a_p'\hat{P}_{-1} + a_uU + dC_{-1}$$

This formulation allows for consumption to respond with an exponentially distributed lag to past labor income, while the response to profits can be expressed as the sum of *two* distributed lags, which allows for the possibility of a longer average lag for this variable. In (8), a_w and a_p measure the short-run *MPC* out of wages and profits, respectively, while the corresponding long-run propensities are given by

$$\hat{a}_w = \frac{a_w}{1-d}; \qquad \hat{a}_\pi = \frac{a_p + a_p'}{1-d}$$

The tests for equation (8) in both linear and ratio form are shown in Table 2 as equations 4*L* and 4*R*.

The results of the linear form confirm that the short-run *MPC* for labor income (around .60) is much larger than that for current profit, around .22. However, the long-run *MPC* out of labor income (.8) is now *lower* than that out of property income which comes to .90. If the coefficients are estimated in ratio form, the results are even more striking: $\hat{a}_w = .8$, $\hat{a}_\pi = 1$. Clearly, these results are entirely inconsistent with the Kaldorian hypothesis—indeed they

---

[11] The estimates of *LD* and *LL* were calculated in full-time equivalent employees. The number of full-time equivalent employees is estimated by adding to the number of full-time employees one-third of so-called "marginal employees," essentially defined as those with less than 32 hours worked per week (even though they would be willing to work longer hours). Source: *Relazione Generale sulla Situazione Economica del Paese.*

TABLE 2—KALDORIAN TESTS: ITALY, 1952–70

	Coefficients of									
	Constant	$\hat{W}$	$\hat{P}$	U	$\hat{P}_{-1}$	C_{-1}	λ	$L(\hat{P}_{-1})$	$S.E.$	$D.W.$
1*L*	1043.4 (4.)	.85 (23.5)	.40 (2.4)						290.5	1.07
1*R*	1350.7 (5.1)	.85	.35 (1.7)						.0142	.80
2*L*	−87.4 (−.2)	.88 (24.6)	.41 (2.7)	.49 (2.1)					263.2	1.45
2*R*	−70 (− .1)	.88	.43 (2.3)	.55 (2.3)					.0127	1.05
3*L*		.88 (33.)	.42 (2.9)	.45 (5.)					255.	1.44
3*R*		.88	.43 (2.4)	.52 (6.3)					.0123	1.04
4*L*		.59 (3.3)	.22 (1.7)	.40 (3.2)	.43 (3.1)	.28 (1.4)			202.9	1.70
4*R*		.55	.23 (1.4)	.44 (3.1)	.45 (2.5)	.31 (1.3)			.0103	1.58
5*L*		.73 (13.1)	.37 (3.1)	.61 (6.6)			.7	1.11 (2.9)	210.8	1.68
5*R*		.68	.41 (3.2)	.68 (9.8)			.7	1.48 (4.2)	.0086	1.77

Note: In the ratio form, the coefficient a_w is obtained by fitting (8) to the data and then adding the coefficient of W/Y to the constant term. For this reason, the regression program does not provide an explicit value of the standard error of a_w, although such a standard error could be computed from the variance-covariance matrix of the errors of the coefficients. We have felt it unnecessary to perform this computation since the t-ratio of a_w is uniformly very large, and of the same order of magnitude as that reported for the linear form. The values in parentheses below the coefficients indicate t-ratios for the corresponding estimates.

are impossible to interpret in that framework.[12]

To be sure, hypothesis (8) is rather restrictive as to possible shapes of the distributed lag for the two components, though the result, implying a very rapid decay for the weights of past labor income, and hump-shaped weights, rising at first then decaying more slowly for $\hat{P}_{-1}$, are certainly consistent with what one might expect a priori. In addition, the previous estimates have one further econometric shortcoming in that the equation is estimated in autoregressive form, which could in principle bias the point estimates of the lagged dependent variables. As an alternative, we assumed that expected labor income could be identified with the current value $\hat{W}$, especially since the possible effects of cyclical variations in employment are taken into account through the unemployment variable. We relied on the specification,

$$(9) \quad C = a_w\hat{W} + a_p\hat{P} + \gamma L(\hat{P}_{-1}) + a_u U$$

in which L is an exponentially distributed lag beginning with $\hat{P}_{-1}$.[13] The rate of decay λ was estimated by scanning over the range $0<\lambda<1$ for the value minimizing the residual variance.[14]

---

[12] The approach of equation (8) was first pursued in a yet unpublished study of Bruno Sitzia. He also pursued some alternative methods of estimating $\hat{a}_r$ all of which confirmed that this coefficient is as large or larger than $\hat{a}_w$. His valuable contribution as well as his advice in designing the tests reported here are gratefully acknowledged.

[13] The Koyck variable for the first year in the sample was obtained by extrapolating backward the positive trend of the profit variable estimated for the first five years (1951–55), until it tended to vanish, and by applying the Koyck transformation to the sample.

[14] As a check, we estimated separate exponentially distributed lags for W; whenever this procedure was tried, the estimate of λ for this component turned out to be around 1 while the estimate of λ for the $\hat{P}$ component was basically unchanged.

From the tests of (9) and of its ratio form reported in equation 5,[15] Table 2, it appears that the distributed lag makes a substantial contribution to the explanation of C; the value of λ in both linear and ratio form indicates a rather slow rate of decay, much as one would expect in view of the short-run variability of profits. The coefficient of current profit is again only half as large as that of labor income. But when we add to it the very high coefficient of lagged profits, we arrive at an estimate of a_p not merely higher than a_w, but actually between 1.5 and 2. This is likely to be a shocking and inadmissible result for those who try to interpret $\hat{a}_\pi$ in the Kaldorian framework as the *MPC* out of property income. As we shall see presently, it is perfectly consistent with the *LCH* framework, which implies an altogether different interpretation.[16] The coefficient of unemployment is again highly significant,[17] though somewhat high, probably reflecting some upward bias.[18]

To summarize, the Kaldorian-type equation of equations 4 and 5 appear to fit quite well the data of the postwar Italian experience. They also yield a sensible estimate of the *MPC* out of labor income and of the effect of structural unemployment. At the same time, the estimated coefficient of profits is of an order of magnitude which is altogether inconsistent with the Kaldorian hypothesis. Hence, if we are to accept this equation as a valid description of saving behavior, it must be on the basis of a framework other than the Kaldorian one. In the following sections, we shall show that such a framework can be provided by the life cycle hypothesis of saving, once the hypothesis is generalized to take into account the stylized facts of development discussed earlier.

IV. The Life Cycle Hypothesis

The life cycle hypothesis (*LCH*), (see Modigliani and Richard Brumberg, 1954), rests on the proposition that the consumption of the representative household reflects the endeavor to achieve the preferred allocation of its life resources to consumption over the life cycle. Together with some further simplifying assumptions, this hypothesis leads to an aggregate consumption function linear in current and expected labor income, YL and YL^e, and in net worth A:

$$\begin{aligned} C &= a'YL + a''YL^e + \delta A \\ &\simeq \alpha YL + \delta A \end{aligned} \tag{10}$$

where α and δ in principle depend on the rate of growth of income and the rate of return on assets.

There are serious problems in attempting to test an equation of the form (10) for Italy since at present there is no available estimate of total consumers' net worth at market prices. The closest proxy available for the period of observation is an estimate of the aggregate value of financial wealth held by households. This estimate,[19] denoted by A^*, falls considerably short of total consumers' net worth A, omitting such items as equity in noncorporate business, and in tangibles directly held by households—residential real estate, land, and the stock of durable goods. Judging from American experience and other very indirect evidence, one might guess that on the average, over the relevant postwar period, A^* has been in the order of one-half to two-thirds of A; but this is no more than an educated guess and, furthermore, this figure might well have fluctuated and also has probably tended to increase in the process of development with the spread of financial intermediation.

---

[15] For the ratio estimate, (9) was transformed to $C/Y = a_p + (a_w - a_p)\hat{W}/Y + \gamma L(P_{-1})/Y + a_u U/Y$.

[16] Similar results were found to hold in the course of the first stage of this research by Tarantelli (1973), even though it was then chosen to give only regression estimates in the neighborhood of the value minimizing the standard error sufficient to reject the Kaldorian hypothesis. An analysis of the results reported here would have implied a framework totally different from the Kaldorian model (see Tarantelli, 1973, note 44, p. 133), and which will be developed in the following pages.

[17] Following similar lines, T. Ter-Minassian, in a yet unpublished study, tested the hypothesis according to which the average propensity to consume for durables should tend to increase as the structural unemployment army decreases. Some preliminary results would also seem to support the symmetrical hypothesis of a fall in the average propensity to consume foodstuffs.

[18] The upward bias may result from our failure to separate the effect of structural and cyclical unemployment, as would be called for by our model. This separation, however, must await the availability of quarterly data, which would provide additional degrees of freedom and thereby also enable us to allow for a subsistence minimum varying as an increasing function of per capita earnings.

[19] Data were based on F. Cotula and M. Caron.

As for the variable YL, since what is conceptually needed is a measure of income from "human capital" which clearly includes not only wages and salaries but also the imputed labor income of farmers and other self-employed (independent) workers, we used the same measure $\hat{W}$, used for the Kaldorian test. Further, in testing the *LCH* equation (10) for a developing country like Italy, we must again allow for the effect of structural and cyclical unemployment. Considerations entirely analogous to those set out in connection with the test of the Kaldorian hypothesis led us to approximate this effect by adding the number of unemployed U to the right-hand side of (10). Thus, the equation finally tested was:

$$C = \alpha\hat{W} + \delta A^* + \beta U \tag{10'}$$

where $\hat{W}$ is defined by equation (7′) and A^* denotes financial wealth deflated, like C and W, by the implicit price deflator for private domestic consumption.

A test of this equation in both linear and ratio form is reported in equations $1L$ and $1R$ of Table 3. The results are rather interesting considering the shortcomings of the data. In the first place, in both cases the equation fits the data appreciably better than any of the "old theories" or, for that matter, the permanent income model tested in Table 1.[20] Second, the coefficient of wealth is highly significant and its value is of the right order of magnitude. The fact that it is somewhat larger than that observed for the United States (.06 ± .02, see Albert Ando and Modigliani, 1963) is to be expected since A^* underestimates private net worth A. In fact, it is probably somewhat on the low side, though this too could be accounted for by the attenuation effect resulting from the "noise" in A^* as a measure of A. Similarly, the coefficient of α is somewhat on the high side but again not by a large factor. One may suspect that the underestimate in the coefficient of A^*, because of attenuation, is absorbed in part by the coefficient of YL and in part by that of U, which is again highly significant but clearly on the high side.[21]

V. The *LCH* versus the Kaldorian Hypothesis

In the test of (10′), as in similar tests for the United States and other countries, it was assumed that the coefficients α and δ could be taken as constant in time. Actually, as mentioned earlier, both coefficients should in principle depend also on r, the rate of return on assets, as was confirmed

---

[20] The reader might wonder to what extent the superiority of the *LCH* results of equation 1, Table 3, over those of the old theories of Table 1 is due to the inclusion of the unemployment variable rather than to the replacement of income with labor income and wealth. To answer this question, the variable U has been added to the equations of Table 1, rows 2 and 3. In both cases, the addition of the unemployment variable improved the fit, but the *S.E.* remained substantially higher than in the life cycle test of row 1, Table 3.

[21] In equation 1 as in all the remaining equations (except 3) in Table 3, we report estimates with the constant term constrained to zero. This specification is consistent with the hypothesis, at least if wealth could be measured directly. Some question about the appropriateness of this specification may, however, arise because of the use of A^* as a proxy for A. In any event, when equation (10′) was estimated with a constant term, this term was uniformly insignificant while the other coefficient estimates hardly changed. For a further discussion of possible bias arising from the use of A^*, see the authors (forthcoming).

TABLE 3—ESTIMATES OF PARAMETERS OF CONSUMPTION FUNCTION: ITALY, 1952–70

		Coefficients of									
	Dependent Variable	YL	P	λ	$L(\hat{P}_{-1})$	P^p	A^*	U	Constant	$S.E.$	$D.W.$
1L	C	.74					.092	1.04		210.	1.19
		(16.)					(4.4)	(6.)			
1R	C/Y	.77					.078	.99		.0113	1.00
							(3.1)	(5.)			
2L	C	.75	.07				.083	1.		216.	1.19
		(14.)	(.4)				(2.7)	(4.6)			
2R	C/Y	.78	.14				.063	.92		.012	.94
			(.6)				(1.8)	(3.9)			
3L	A^*		4.75	.7	11.9	16.6			− 9.374	2648.	.77
			(3.2)		(5.9)				(− 5.1)		
3R	A^*/Y		6.2	.6	10.3	16.5			−11.208	.118	.86
			(4.)		(5.1)				(− 7.4)		
4L	C	.64	.13	.8	1.21	1.34	.060	.97		191.	1.60
		(9.2)	(.8)		(2.3)		(2.1)	(5.1)			
4R	C/Y	.59	.28	.8	1.9	2.18	.03	.85		.0085	1.86
			(1.6)		(3.7)		(1.1)	(4.9)			

Note: The values in parentheses below the coefficients are t-ratios.

by some numerical calculations in the unpublished paper by Modigliani and Brumberg, reported in Ando and Modigliani (1963). However, the nature of this dependence is rather complex. The difficulty here is the familiar one that a rise in r (which should be interpreted here as the long-term *real* rate of return, since variations in the short-term rate around a stable long-run level would have negligible effects), labor income constant, affects the life profile of consumption both through an income effect, tending to raise consumption at all ages, and a substitution effect tending to shift consumption away from the present and toward a later age. In the case of current consumption these two effects go against each other and the net effect cannot be ascertained unless one is prepared to make specific assumptions about the nature of the utility function. The numerical results cited above were obtained for the limiting case in which the substitution effect was assumed zero (corner indifference curves or strict complementarity). As expected, they showed that an increase in r would in steady state lead to a fall in the wealth-income ratio A/Y, and thus to a fall in the saving ratio $S/Y=\rho A/Y$, for any given rate of growth ρ. In terms of the consumption function (10), it was found that a rise in r tended to raise both α and δ. The effect on α was relatively minor but on δ it was proportionately quite large. Specifically, δ appeared to increase linearly with r, or $\delta(r)=\delta^*+\mu r$, with μ roughly unity (and δ^* around .08).

These results of the unpublished paper have since led Modigliani (1973) to conjecture that the effect of r on the LCH consumption function might be approximated by rewriting (10) in the form,

$$\begin{aligned}(11)\qquad C_t &= \alpha YL_t + (\delta^* + \mu r_t)A_t \\ &= \alpha YL_t + \delta^* A_t + \mu(r_t A_t)\end{aligned}$$

with the value of μ depending on the substitution effect between current and future consumption.[22] The value of μ in turn is critical in determining the steady-state response of the saving ratio to changes in r. Indeed, as shown in Modigliani (1973), equation (11) implies the following steady-state relation between the saving ratio, μ, r, and ρ.

---

[22] An expression analogous to (11) has also been derived as a linear approximation by W. H. Somermeyer and R. Bannink, starting from an additive logarithmic utility function except that α is also replaced by α_0+

$$\frac{S}{Y} = \frac{\rho(1 - \alpha)}{\delta^* + \rho + (\mu - \alpha)r} \tag{11'}$$

The calculations of Modigliani-Brumberg reported above suggest that μ might have an upper bound of one in the limiting case of no substitution. Since $\alpha < 1$, it is seen from (11′) that such a value of μ implies that a rise in r reduces S/Y, as expected. But μ would be lower in the presence of substitution. Some simulations reported by James Tobin and Walter Dolde, assuming an iso-elastic additive utility function with an elasticity of 1.5, suggest a value of μ close to zero (see Table 3, simulation 13), which implies that S/Y is an increasing function of r. Another set of calculations reported by Tobin (1967), assuming a logarithmic utility function, implying an elasticity of one and hence greater substitution, shows a stronger positive responsiveness of S/Y to r, suggesting a negative value of μ. Unfortunately, there is no way to establish a priori just how strong the substitution effect might be in reality, and hence what value μ should take—except that it should not exceed unity. Because of a hunch that the substitution effect is unlikely to be strong relative to the income effect, we would rather expect μ positive and not very different from α (which, as can be seen from (11′) would imply that S/Y is roughly independent of r). However, in principle, it could be a good deal lower and even negative.

The generalization of the *LCH* embodied in equation (11) is likely to be especially useful for a developing economy, since the process of development might tend to be accompanied by systematic changes in the rate of return on assets. In addition, it provides a useful link to the Kaldorian form of the consumption function. This can best be seen by noting that r_tA_t may be thought of as the long-run expectation of the flow of profits out of current assets, call it P^p, the capitalization of which at the long-term rate r determines the market value of wealth, i.e., $A_t = P_t^p/r_t$. Hence, (11) can be rewritten in the form

$$C_t = \alpha YL_t + \overset{*}{\delta} A_t + \mu P_t^p \tag{12}$$

Comparing (12) with (4), it is apparent that there are some important similarities between *LCH* and *KH* in that both models assert that C is a linear homogeneous function of labor income and profits, and with different coefficients. The main differences are:

(i) In terms of the arguments in the consumption function: For *LCH*, they are the expected rate of earnings from human effort, closely related to Friedman's return from human capital, and the expected long-run return from current private wealth, in addition to the value of wealth itself. In *KH*, the last variable does not appear and the first two variables supposedly denote current "wages" and "profits"—or, at any rate, no distinction is made between current measured values and long-run expectations.

(ii) In terms of interpretation: For *KH*, the coefficients of YL and P are different because YL and P are received by the members of different castes. The coefficients of these variables can rightly be interpreted as marginal propensities to consume of the members of each caste. In the *LCH*, on the other hand, every household typically earns both kinds of income during at least part of its life cycle, and the coefficient of P^p *cannot* be interpreted

---

$\alpha_1 r_t$. They also show that μ should be negative, but this result reflects the special assumption of a logarithmic utility function. See text below.

as a marginal propensity. It measures, instead, the strength of the substitution effect between current and future consumption. The *LCH* per se has little to say about the value of this coefficient or even about its sign.

Yet the *LCH* permits some inferences about the results that one might expect to find if one were to estimate the coefficient of P^* in the Kaldorian equation (4). Consider first the quantity dC/dP^p which measures the response of C to a change in expected property income. Differentiating (12) totally with respect to P^p, and remembering the relation $P^p - rA = 0$, one obtains

$$\frac{dC}{dP^p} = \mu + \frac{\delta^*}{r} E_{AP} \equiv c_p \tag{13}$$

where E_{AP} is the elasticity of A with respect to P^p. Thus, according to the *LCH* the response of consumption depends on the nature of the change in P^p. If it changes because of a change in A, r constant—which is presumably the most common source of change—then E_{AP} will be one and $dC/dP^p = \mu + \delta^*/r$. But if it changes only because of a change in r, then E_{AP} is zero and $dC/dP^p = \mu$. Thus an increase in profit due to a change in A is sure to increase C and the derivative is likely to exceed α; but, if due to a change in r, it might well be less than α and it could be even negative.

If, as in the present case, the observations are time-series of aggregates for a given country, presumably most of the observed variation in profits must come from variations in A (through saving and dissaving) rather than from variations in r. Hence c_p of (13) must be close to $\mu + \delta^*/r$. Now, as we have seen, the value of δ^* seems to be in the order of .07 and hence δ^*/r is certainly positive and of the order of unity—and could easily be above one. On the other hand, the value of μ is quite uncertain: it is unlikely to exceed unity, but it could be smaller and even negative. Thus c_p could be larger than α and, in fact, it *might well exceed unity*. However, c_p is the coefficient of P^p. On the other hand, in testing the Kaldorian model, it has been customary to use *current* labor income and *current* profit. The difference is not too important in the case of YL because labor income is quite stable; hence, expected income will tend to be roughly proportional to current YL. But the situation is entirely different in the case of P. As was pointed out by Ando and Modigliani (1963), since P tends to fluctuate substantially in response to fluctuations in aggregate demand, P^p must be based on an average experience from the past (as well as other information). Hence, $dC/dP = c_p dP^p/dP$ will be distinctly smaller than c_p, since dP^p/dP is certainly smaller than one and may be close to zero if P exhibits much short-run variation. Thus, independently of the relation of c_p to α, a regression of C on YL and P is most likely to yield a P coefficient much smaller than that of YL, especially since YL will tend to proxy in part for P^p. In the case of the United States, for example, Ando and Modigliani (1963) obtained a coefficient of YL of .93 and a coefficient of P of .07 (totally insignificant). This was true even though A, in contrast to P, contributed very significantly to the explanation of C.

The results of a similar test for Italy have already been reported in connection with our test ~~of the~~ Kaldorian hypothesis and appear in Table 2, equations 2*L* and 2*R*. From those results, one can infer that the coefficient of current profit is only .43, and hence much lower than the coefficient of wages. But according to *LCH*, this low coefficient does not reflect a purportedly low *MPC* out of property income, but rather the fact that current profit P is a poor proxy for wealth and/or expected long-run profit P^p, the truly relevant variables. The supposition receives some indirect support from the fact that the coefficient of P is not only low but also has a

t-ratio which is relatively low, especially when compared with that of $\hat{W}$ (see equation 2) or with that of A in equation 1, Table 3. These results are thus in line with those of Ando-Modigliani (1963) for the United States. But a more telling test, reported in equations $2L$ and $2R$ of Table 3, consists in adding the variable A to the Kaldorian model (or equivalently P to the LCH model); in the linear estimate, the coefficient of P then falls from .42 to .07 and is insignificantly different from zero whereas the A coefficient falls only from .092 to .083 and is still quite significant. Similar results hold for the ratio estimate. These results are consistent with the LCH interpretation that P, in the Kaldorian equations 2 of Table 2, is but a poor proxy for the relevant variables A and P^p.

The validity of this inference is supported by the other results which are reported in Table 2. Indeed, we can now interpret these tests as an endeavor to estimate the coefficient c_p of equation (13), by replacing both A^* and P by an approximation to P^p. As we have already seen in commenting on the results of Table 2, we find that the coefficient of P^p is uniformly larger than that of wages, and that in fact most of the time our estimate is above unity. In particular, when using the more refined approach of equation 5, relying on a Koyck distributed lag on $\hat{P}_{-1}$, our estimate of c_p is as high as 1.5 to 1.9, a value which is perfectly consistent with the LCH framework, as is apparent from (13). This consistency can also be verified by relating the estimate of c_p to the estimate of the coefficient of A^*, δ, in equation 1, Table 3. From equations (10), (11), and (13) it is seen that δ, and the coefficient of P^p, c_p, should be related by

$$(14) \qquad c_p \simeq \frac{\delta^*}{\bar{r}} + \mu \equiv \frac{1}{\bar{r}}(\delta^* + \mu\bar{r}) = \frac{1}{\bar{r}}\delta$$

where $\bar{r}$ is the average value of the return on wealth, $r = P^p/A$ or equivalently the capitalization rate relating the market value of wealth to expected profits, $A = P^p/r$.

Relying on the assumption that r was reasonably stable in the relevant period and that, as in Table 2, P^p can be approximated by a linear combination of $\hat{P}$ and a Koyck lag beginning with $\hat{P}_{-1}$,[23] we can obtain an estimate r, $\hat{r}$, from the regression

$$(15) \qquad A^* = \mu'\hat{P} + \mu''L(\hat{P}_{-1}) + a$$

where $\mu' + \mu'' \simeq 1/\hat{r}$. The results obtained in estimating (15) are shown in equations $3L$ and $3R$, Table 3. The fit is reasonably good considering the rough approximation to A. The estimate of $1/\bar{r}$ implied by $\mu' + \mu''$, 16.5, is not unreasonable (though somewhat on the high side).[24] It is seen from (15) that this estimate of $1/r$ together with the estimate of δ of Table 3 would lead us to expect a value of c_p of around 1.5 (.092x16.6 relying on the linear estimates, .078x16.5 for the ratio estimates), which is quite consistent with the actual estimate.[25]

An attempt was also made to estimate the coefficients of the LCH function in its more general form (12), which includes both A and P^p, in the hope of gaining some

---

[23] We allow for a constant term on the ground that what we measure is not A, but financial wealth A^*, and we hypothesize that the relation of A^* to A can be approximated by

$$A^* = a_0A - a_k = \frac{a_0}{r}P^p - a_1; \quad 0 < a_0 < 1, a_1 > 0$$

since, in the process of development, financial wealth is likely to rise proportionately faster than A (see fn. 21).

[24] Especially since from fn. 23, it is seen that $\mu_1 + \mu_2$ is an estimate of a_0/r and $a_0 < 1$.

[25] Since in the test equations 1 and 2, A^* denotes the value of assets at the end of the *preceding* year, the same is true in (15), and equation 3 of Table 3. Nonetheless, the explanatory variables include *current* profits $\hat{P}$. This was done on the ground that the variable $\hat{P}$ appears in equation 5 of Table 2 and the estimate of c_p is the sum of the coefficient of profits including the current value $\hat{P}$. The very significant contribution of current $\hat{P}$ to the explanation of earlier A^* can be readily accounted for on the hypothesis that $\hat{P}$ is a good proxy for profit expectations for the coming year held at the end of the previous year, which are reflected in A^*.

information on the sign and size of the coefficient of P^p, an estimate of μ in (12) which reflects the strength of the substitution effect.

This attempt runs into serious difficulties because of the unavoidable collinearity of A and P^p, a problem made more serious in the present instance by our use of A^* as a proxy for A. The results are reported in equation 4, Table 3, although we regard them as rather unreliable. They do suggest that the value of μ is positive; however, the point estimate, greater than unity, is above the a priori upper limit, probably as a result of upward biases arising from the fact that our approximation to A is very rough, and hence P^p may partly act as a proxy for A.

On the whole, and subject to the caveat at the beginning of the previous paragraph, the results of equation 4, Table 3 seem to warrant the conclusion that a generalized version of the life cycle model can provide an explanation of consumption in postwar Italy which is substantially superior to that offered by any of the competing hypotheses. They also confirm and throw light on the inference that c_p is well above one. They further suggest that μ of equation (12) is positive at least in the case of Italy, though its numerical value (and therefore, in particular, its relation to α) cannot be reliably estimated with the presently available data.

While these striking results for Italy have not yet been replicated, they receive some support from scattered evidence for other countries. First, Richard Stone, in his 1966 paper, provides us with an explicit estimate of δ^* and μ of equation (12) for the United Kingdom, though his results are based on some rather stringent assumptions which may limit their reliability for our purpose.[26] His estimate of α is .79 and that of μ is positive and in fact only moderately lower, .67; at the same time, the separation of income into labor and property income does not affect the estimate of δ^* which remains .06. It is readily apparent from these estimates that for any reasonable value of the rate of return r, the quantity $.67+.06/r$ will exceed the estimated *MPC* out of labor income of .79, and in fact will exceed unity. These results have been confirmed by his later paper extending the analysis to 1970. In particular, when relying on the alternative measure of consumption excluding durables, the estimates of both α and μ are reduced as expected, but, in addition, the difference between them shrinks further to only about .10 (and with a very small t-ratio).

Finally, for the United States, Warren Weber, following a rather different method of estimation, but allowing explicitly for the effect of the interest rate, has concluded that the partial derivative of C with respect to r is positive, implying again a positive value of μ and, hence again, a value of c_p most probably above unity or at least above α.

One suggestive implication of these conclusions should be to relieve the schizophrenia of many economists and other social scientists who have been torn in the past between advocating a redistribution of income in favor of labor on grounds of equity, and reluctantly accepting a redistribution in favor of profits on the ground that this would increase savings and capital formation and, hence, the potential rate of economic development. Our evidence suggests that a sheer redistribution of a given income from wages to profits would increase savings only in the short-run, but would decrease it in the longer run (since c_p is larger than α). Furthermore, a reduction in the rate of return or rate of profit resulting from a lower productivity of capital or from taxation of property income might also increase saving, although this inference is uncertain since the available evidence does not enable us to conclude with confidence that μ exceeds α in (12), (see equation (11′)).

---

[26] The most sensitive assumptions for our purpose are that: (i) permanent labor income can be approximated by a Koyck lag of the wage and salary component of National Income; (ii) permanent property income P^p is a Koyck distributed lag of the remaining components of income; and (iii) the weights of the distributed lag are the same for the two components.

REFERENCES

C. Ackley, "The Wealth-Saving Relationship," *J. Polit. Econ.*, Feb./Dec. 1951, *59*, 155–61.

A. Ando and F. Modigliani, "The 'Life Cycle' Hypothesis of Saving: Aggregate Implications and Tests," and "Correction," *Amer. Econ. Rev.*, Mar. 1963, *53*, 55–84 and Mar. 1964, *54*, 111–13.

D. S. Brady and R. D. Friedman, "Savings and the Income Distribution," Nat. Bur. Econ. Res. *Stud. in Income and Wealth*, Vol. 10, New York 1947, 247–65.

T. M. Brown, "Habit Persistence and Lags in Consumer Behavior," *Econometrica*, July 1952, *20*, 355–71.

R. E. Brumberg, "An Approximation to the Aggregate Saving Function," *Econ. J.*, May 1956, *66*, 66–72.

G. Carlucci and G. M. Rey, "La funzione aggregata del consumo in Italia," *Riv. Polit. Econ.*, Nov. 1963, *53*, 3–23.

F. Cotula and M. Caron, "I conti finanziari dell'Italia. Dimensioni e struttura della riccheza e del risparmio finanziari dell'economia," Banca d'Italia *Bollettino*, Nov.-Dec. 1971, *6*, 839–919.

T. E. Davis, "The Consumption Function as a Tool for Prediction," *Rev. Econ. Statist.*, Aug. 1952, *34*, 270–77.

F. Di Fenizio, "La funzione del consumo in Italia come legge di lungo periodo," *L'Industria*, Jan.-Feb. 1958, *1*, 37–59.

J. S. Duesenberry, *Income, Saving and the Theory of Consumer Behavior*, Cambridge, Mass. 1949.

M. J. Farrell, "The New Theories of the Consumption Function," *Econ. J.*, Dec. 1959, *69*, 678–96.

I. Fisher, *The Theory of Interest*, New York 1930.

M. Friedman, *A Theory of the Consumption Function*, Princeton 1957.

R. W. Goldsmith, *A Study of Saving in the United States*, Princeton 1956.

A. Graziani, *Reddito Nazionale, Moneta e Consumi Nell'economia Italiana: Saggi*, Naples 1961.

N. Kaldor, "Alternative Theories of Distribution," in his *Essays in Value and Distribution*, London 1960.

J. M. Keynes, *The General Theory of Employment, Interest and Money*, New York and London 1936.

L. M. Koyck, *Distributed Lags and Investment Analysis*, Amsterdam 1954.

S. Kuznets, *Uses of National Income in Peace and War*, Occas. Paper 6, Nat. Bur. Econ. Res., New York 1942.

F. Modigliani, "Fluctuations in the Saving-Income Ratio: A Problem in Economic Forecasting," Nat. Bur. Econ. Res. *Stud. in Income and Wealth*, Vol. 11, New York 1949.

———, "The Life Cycle Hypothesis of Saving, the Demand for Wealth and the Supply of Capital," *Soc. Res.*, Summer 1966, *33*, 160–217.

———, "The Life Cycle Hypothesis of Saving and Inter-Country Differences in the Saving Ratio," in W. A. Eltis et al., eds., *Induction, Growth and Trade: Essays in Honour of Sir Roy Harrod*, Oxford 1970, 197–225.

———, "The Life Cycle Hypothesis of Saving Twenty Years Later," *Proc. AUTE Conference*, Univ. Warwick, England, Mar. 1973.

——— and A. Ando, "The Permanent Income and the Life Cycle Hypothesis of Saving Behavior: Comparisons and Tests," *Consumption and Savings*, Vol. 2, Philadelphia 1960, pp. 156–59.

——— and ———, "Tests of the Life Cycle Hypothesis of Saving," *Bull. Oxford Univ. Inst. Econ. Statist.*, May 1957, *19*, 99–124.

——— and R. Brumberg, "Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data," in K. Kurihara, ed., *Post Keynesian Economics*, New Brunswick 1954.

——— and ———, "Utility Analysis and Aggregate Consumption Function: An Attempt at Integration," unpublished paper, 1954.

——— and E. Tarantelli, "A Generalization of the Phillips Curve for a Developing Country," *Rev. Econ. Stud.*, Apr. 1973, *40,* 203–23.

——— and ———, *Mercato del Lavoro, Distribuzione del Reddito e Consumi Privati,* il Mulino, Bologna (forthcoming).

A. Niccoli, "Consumi privati," in *Analisi quantitativa per la programmazione di breve periodo,* Istituto Nazionale per lo Studio della Congiuntura (ISCO), Rome 1971.

D. Ricardo, "On the Principles of Political Economy and Taxation," in P. Sraffa, ed., *Works and Correspondence of David Ricardo, 1,* Cambridge 1953, 96–97.

B. Sitzia, "A Note on Consumption Functions in Econometric Models," unpublished study, Rome 1974.

W. H. Somermeyer and R. Bannink, *A Consumption-Savings Model and Its Applications,* Amsterdam and London 1973.

R. Stone, "Spending and Saving in Relation to Income and Wealth," *L'Industria,* Oct.-Dec. 1966, *4,* 471–99.

———, "Personal Spending and Saving in Postwar Britain," in W. Sellekaerts, ed., *Economic Development and Planning: Essays in Honour of J. Tinbergen,* White Plains 1974.

P. Sylos-Labini, "Prices, Distribution and Investment in Italy, 1951–1966: An Interpretation," *Banca Naz. Lavoro Quart. Rev.*, Dec. 1967, *23,* 316–70.

———, "Sviluppo economico e classi sociali in Italia," *Quaderni di Sociologia,* Oct.-Dec. 1972, *4,* 371–443.

E. Tarantelli, "Produttività del lavoro, salari e inflazione," Ente per gli Studi Monetaria, Bancari e Finanziari L. Einàudi, *Quaderni di Richerche,* Jan. 1970, *5,* 1–128.

———, (1973a) "Note sul consumo nella teoria economica e nuove linee di ricerca per l'esperienza italiana," *Riv. Int. Sci. Soc.*, Jan.-Apr. 1973, *81,* 115–33.

———, (1973b) "Distribuzione del reddito e rinnovi contrattuali nell'esperienza italiana," *Riv. Int. Sci. Soc.*, July-Oct. 1973, *81,* 345–72.

T. Ter-Minassian, "Effects of Tax Changes on the Real Variables of the Economic System in Italy," unpublished study, Rome 1973.

J. Tobin, "Life Cycle Saving and Balanced Growth," in *Ten Economic Studies in the Tradition of Irving Fisher,* New York 1967.

———, "Asset Holdings and Spending Decisions," *Amer. Econ. Rev. Proc.*, May 1952, *42,* 109–37.

——— and W. C. Dolde, "Wealth, Liquidity and Consumption," in *Consumer Spending and Monetary Policy: The Linkages,* Fed. Reserve Bank Boston, Monetary Conference Series, No. 5, Boston 1971.

A. Ulizzi, "Income, Saving and Expenditure of Italian Households in 1968: Review of the Economic Conditions in Italy," *Banco di Roma,* Nov. 1970, *24,* 410–38.

S. Vinci, "Consumi e distribuzione del reddito," *L'Industria,* Jan.-Mar. 1971, *1,* 21–39.

W. R. Weber, "The Effect of Interest Rates on Aggregate Consumption," *Amer. Econ. Rev.*, Sept. 1970, *60,* 591–600.

Banca d'Italia, *Relazione Annuale,* Rome 1971.

———, Gruppo per lo Studio della Politica Monetaria e Fiscale, "Un modello econometrico dell'economia italiana M.2B.1," mimeo., Rome 1974.

Ministero del Bilancio, *Relazione Generale sulla Situazione Economica del Paese,* annual issues, 1950–70.

Interest Rates, Inflation, and Consumer Expenditures

By Warren E. Weber*

The possibility that consumers use different decision rules in making different types of expenditure decisions has been explored in several recent econometric models which use multiequation models to explain consumer expenditures. Two such models are the Brookings model and one particular vintage of the Federal Reserve Board–Massachusetts Institute of Technology (*FRB-MIT*) model.[1] However, these two models do not provide any microeconomic foundations for their regression equations. This neglect leads Daniel Suits and Gordon Sparks to leave interest rates and relative price effects out of the automobile and consumer durables demand equations in the Brookings model. It also leads the *FRB-MIT* model builders to leave interest rates and relative prices out of the determination of the level of consumer expenditures. The only role of interest rates and relative prices is to allocate consumer expenditures among their various components.[2]

The purpose of this paper is to develop a multiequation model of consumer expenditures which uses the individual's intertemporal utility maximization calculus to explain his purchases of durables and nondurables. The approach will extend the basic Fisherine analysis to the case in which the individual can purchase these two types of commodities. For earlier applications, see the consumption function studies of Franco Modigliani and Richard Brumberg, Albert Ando and Modigliani, Milton Friedman, and the author. In addition to human and nonhuman wealth (which includes existing stocks of durables), the consumer demand equations obtained have a nominal interest rate, the relative price of the two commodities, and the anticipated rates of change in the prices of the two commodities as exogenous variables. The construction of the model is such that the level of consumer expenditures may be affected by all of these exogenous variables. Moreover, the nominal interest rate, the relative price of the two commodities, and the anticipated rate of change in the price of durables may affect the allocation of expenditures. These demand functions are then used to test the statistical significance of the interest rate and price variables as determinants of consumer expenditures.[3]

The paper proceeds as follows. Section I contains the derivation of the demand functions for durables and nondurables of a representative consumer. In Section II, specific functional forms for these demand functions are obtained by assuming a specific functional form for the representative consumer's utility function. Section III presents the results of estimating the demand functions derived in Section II using aggregate annual data for the United States for the period 1930–70, omitting 1941–46. Section IV contains a summary and conclusions.

---

* Associate professor of economics, Virginia Polytechnic Institute and State University.

[1] Specifically I refer to the vintage described by Frank de Leeuw and Edwin Gramlich.

[2] This part of the *FRB-MIT* model is described by de Leeuw and Gramlich as follows:

> The sum of the services of durable goods and expenditures on non-durables is the consumption variable that we relate to current and past income, whereas the allocation of the sum among its components depends on relative prices, existing stocks, and other variables. One result of this formulation is a small effect of interest rates on the allocation of total consumption (in our sense) and hence on consumer expenditures on durable goods. [p. 12]

[3] Another study which develops simultaneous expenditure functions is that by W. E. Diewert. The major difference is that the present study explicitly tests the statistical significance of interest rates, relative prices, and anticipated inflation rates as determinants of expenditures.

I. The Representative Consumer's Utility Maximization Problem

Suppose that a representative consumer in the economy begins his economic life at age 0 and lives L years with certainty. He is currently assumed to be at economic age i, so that he has a remaining planning horizon of $L-i-1$ years. He is assumed to receive utility from his consumption of the services of nondurable goods and the services of durable goods in each of the years remaining in his economic lifetime. We assume that the service flows from nondurables and durables are not perfect substitutes in his utility function. Let the services of nondurables and durables which he consumes at age j be denoted by s_j^n and s_j^d, respectively. Under the further assumption that he will not receive any utility from bequests, we can express his utility index as

$$(1)\quad U = U(\overset{n}{s}_i, \ldots, \overset{n}{s}_{L-1}, \overset{d}{s}_i, \ldots, \overset{d}{s}_{L-1})$$

Three further assumptions concerning the representative consumer are now made: 1) He receives an after-tax, nominal labor income of y_j at each age $j=i, \ldots, L-1$. 2) His only alternative to purchasing the services of nondurables or the services of durables in any period is to purchase one-period bonds which promise to pay one dollar at the beginning of the next period. 3) He owns no durables. Under these assumptions his multiperiod wealth constraint is[4]

$$(2)\quad \sum_{j=i}^{L-1}\left(\frac{1}{1+r}\right)^{j-i}(\overset{n}{q}_j\overset{n}{s}_j + \overset{d}{q}_j\overset{d}{s}_j) = b_i + \sum_{j=i}^{L-1}\left(\frac{1}{1+r}\right)^{j-i} y_j$$

where r is the nominal interest rate, b_i is his bond holdings at the beginning of period i, and q_j^n and q_j^d are the prices of nondurable services and durable services which he faces at age j. The multiperiod wealth constraint states that the present value of the individual's consumption of the services of nondurables and durables must equal his total human and nonhuman wealth.

Now define v_i the individual's total of human and nonhuman wealth, and $\underline{q}_i^n$, $\underline{q}_i^d$ the vectors of forward nondurable and durable service prices he faces at age i:

$$(3\text{a})\quad v_i = \sum_{j=i}^{L-1}\left(\frac{1}{1+r}\right)^{j-i} y_j + b_i$$

$$(3\text{b})\quad \underline{\overset{n}{q}}_i = \left\{\overset{n}{q}_i, \left(\frac{1}{1+r}\right)\overset{n}{q}_{i+1}, \ldots, \left(\frac{1}{1+r}\right)^{L-i-1}\overset{n}{q}_{L-1}\right\}$$

$$(3\text{c})\quad \underline{\overset{d}{q}}_i = \left\{\overset{d}{q}_i, \left(\frac{1}{1+r}\right)\overset{d}{q}_{i+1}, \ldots, \left(\frac{1}{1+r}\right)^{L-i-1}\overset{d}{q}_{L-1}\right\}$$

Then, when he maximizes his utility function (1) subject to the wealth constraint (2), his $2(L-i-1)$ demand functions for the services of nondurables and durables will be:

$$(4\text{a})\quad \overset{n}{s}_j = \overset{n}{s}_j(\underline{\overset{n}{q}}_i, \underline{\overset{d}{q}}_i, v_i)\qquad j=i, \ldots, L-i-1$$

$$(4\text{b})\quad \overset{d}{s}_j = \overset{d}{s}_j(\underline{\overset{n}{q}}_i, \underline{\overset{d}{q}}_i, v_i)\qquad j=i, \ldots, L-i-1$$

These demand functions are homogeneous

---

[4] This wealth constraint also contains the implicit assumption that the individual can borrow and lend at the same interest rate in all periods. This assumption is made to simplify the empirical analysis of this paper.

of degree zero in q_i^n, q_i^d, and v_i.

The functions (4a) and (4b) are the representative consumer's demand functions for services of nondurables and durables. They would be sufficient to explain consumer expenditures on durables and nondurables if he were able to rent the services of these two commodities directly. In general, however, this is not the case for nondurables since usually units of the commodity must be purchased in order to obtain the service flow. Further, we will assume that a rental market for the services of durables does not exist and that the only way in which the representative consumer can obtain the services of durables is to purchase units of the durable good and then "rent" the service to himself.[5]

Thus, there are two problems in converting the service demand functions of the representative consumer into the purchase demand functions for units of nondurables and durables. The first is determining the service flows which are obtainable from newly purchased and existing units of the commodity. The second, which arises here only for durables, is determining the nature of the resale market for existing units of the good. Of course, the relationship between the service flows obtainable from newly purchased units of the commodity and existing units of the commodity may have definite implications for the resale market. Thus, the solution of the first problem may have definite implications for the solution of the second problem.

We solve the service flow and rental market problems for nondurables as follows. The service flow from units of nondurables is assumed to be equal to the units of nondurables purchased during the period. Since nondurables in our model are goods which fully depreciate by the end of the period, there is no resale market to consider. Consequently, a demand for the services of nondurables causes an immediate demand to purchase nondurables.

The service flow and rental market problems for durables are handled in the following way. The service flow from a particular stock of durables is assumed to be a constant proportion of the size of this stock. In addition, the service flow from a unit of existing durables *properly depreciated* is assumed to be same as that from a unit of newly purchased durables. For example, a one-year old TV is assumed to yield a service flow which is $(1-\delta)$ times the service flow for a new TV, where $0<\delta<1$ is the proper depreciation rate.[6] Combining these two assumptions means that the flow of durable services is a constant proportion of the stock of durables held in the period, when this stock is measured in units of new durables.

These assumptions concerning service flows from new and existing durables have definite implications for the resale market for existing durables. In the absence of transactions costs in the sale of existing durables, these assumptions imply that existing durables and new durables must, by the law of indifference, sell at the same price per unit of remaining potential service. These assumptions also imply that a demand for the services of durables gives rise to a reservation demand for durables. However, unlike the demand for nondurable services, the demand for durable services will give rise to a purchase demand for durables only if the reservation demand for durables differs from the individual's current holdings of durables.

---

[5] This assumption does not appear unreasonable when it is considered that our representative consumer's durable purchase demand function will be our aggregate demand function. However, it does point out a double-counting problem which would arise if interhousehold rentals of durables were included in the data used in the empirical analysis. This point will be considered later in the paper.

[6] There are, of course, other depreciation assumptions which could be made, such as the "one-horse shay" type mentioned by Richard Parks. An alternative type of assumption would be that the depreciation rate depends upon the usage of the durable. However, we choose the constant, use-independent depreciation rate here for convenience.

Let n_j be the consumer's purchases of nondurables at age j and let d_j be his *holdings* of durables during age j. Then under our assumptions relating the flows of services of nondurables and durables to the stocks of these commodities, we can replace s_j^n by n_j and s_j^d by d_j in the individual's utility function, and his utility function can now be written as

$$(1') \quad U^* = U^*(n_i, \ldots, n_{L-1}, d_i, \ldots, d_{L-1})$$

Further, his intertemporal decision problem becomes the maximization of (1′), subject to the wealth constraint

$$(2') \quad v_i = \sum_{j=i}^{L-1} \left(\frac{1}{1+r}\right)^{j-i} \cdot \left\{ p_j^n n_j + p_j^d [d_j - (1-\delta) d_{j-1}] \right\}$$

where p_j^n and p_j^d are the purchase prices for nondurables and durables, respectively, which the individual faces at age j.

This wealth constraint, like the wealth constraint (2), is derived under the implicit assumption that the representative consumer believes that he will be able to fully transact his demands in all periods. In other words, the representative consumer is assumed not to face any quantity restrictions on his ability to purchase nondurables at the prevailing prices, on his ability to purchase or sell bonds at the prevailing prices, or on his ability to adjust his durable holdings to his desires. In order for the individual to encounter no quantity adjustment difficulties with respect to durables, we require a well-functioning resale market for existing durables and a perfectly elastic supply curve of new durables at the prevailing price.[7]

From the wealth constraint (2′) we find that we must redefine the individual's total human and nonhuman wealth to include the durables which he owns coming into period i. Further, we find that we can now define the vectors of forward service prices in terms of current and future purchase prices. We will assume throughout that the proportion of service flows to units for both nondurables and durables is unity. Thus we can rewrite (3a)–(3c) as[8]

$$(3a') \quad w_i = v_i + (1-\delta) p_i^d d_{i-1}$$

$$(3b') \quad \underline{p}_i^n = \left\{ p_i^n, \left(\frac{1}{1+r}\right) p_{i+1}^n, \ldots, \left(\frac{1}{1+r}\right)^{L-i-1} p_{L-1}^n \right\}$$

$$(3c') \quad \underline{p}_i^d = \left\{ p_i^d - p_{i+1}^d \left(\frac{1-\delta}{1+r}\right), \left(\frac{1}{1+r}\right) \cdot \left[p_{i+1}^d - p_{i+2}^d \left(\frac{1-\delta}{1+r}\right) \right], \ldots, \left(\frac{1}{1+r}\right)^{L-i-1} \cdot \left[p_{L-1}^d - p_L^d \left(\frac{1-\delta}{1+r}\right) \right] \right\}$$

Then, when the individual maximizes his utility function (1′) subject to the wealth constraint (3a′), his $2(L-i-1)$ demand functions are

$$(4a') \quad n_j = n_j(\underline{p}_i^n, \underline{p}_i^d, w_i) \quad j = i, \ldots, L-1$$

---

[7] It may be objected that whereas a well-developed market for the services of some types of durables does exist, a well-developed resale market only exists for automobiles and some types of appliances. Therefore, a more reasonable assumption would be that an upper bound on the decrease in durables holdings between two periods should be the depreciation rate on durables. We ignore this problem because our demand functions for the representative consumer will be our aggregate demand functions and for the time period under consideration we never observe zero purchases of durables.

[8] The redefinition of q_{L-i-1}^d given by (3c′) is not strictly correct. Since the representative consumer will not be alive in period L, the price of services of durables in $L-i-1$ is $q_{L-i-1} = p_{L-i-1}$. However, since we will assume a long-lived consumer, we will ignore this point here and in the empirical analysis.

$$(4b') \quad d_j = d_j(\underline{p}_i^n, \underline{p}_i^d, w_i) \quad j = i, \ldots, L-1$$

These demand functions are homogeneous of degree zero in $\underline{p}_i^n$, $\underline{p}_i^d$, and w_i.

II. The Specific Model

A. *The Explicit Demand Functions*

In the previous section we developed a prototype model of the consumer's lifetime allocation process for nondurables and durables. In this section we transform the implicit demand functions of the prototype model into explicit demand functions suitable for empirical estimation.

One possible approach to this problem would be to assume that (4a′) and (4b′) are either linear or log-linear in their arguments. However, this solution presents the following difficulty. The classical utility maximization calculus implies certain restrictions on the partial derivatives of a set of demand functions. In our case, however, we are not able to observe the entire set of demand functions since the demands for purchases of nondurables and durables in future periods are not observable. Therefore, in the empirical analysis we would not be able to take advantage of the additional information conveyed by these restrictions.

Thus, our approach to the problem will be to assume that the representative consumer has a utility function of a specific functional form. Maximizing this utility function subject to the consumer's intertemporal wealth constraint will yield explicit demand functions for current purchases of nondurables and durables which place some restrictions on the partial derivatives of these two demand functions.[9]

We first assume that the consumer's utility index for nondurable and durable holdings during any age j is a *CES* utility function of the form

$$(5) \quad u(n_j, d_j) = [\alpha n_j^{-\beta_1} + (1-\alpha) d_j^{-\beta_1}]^{-1/\beta_1} \qquad j = i, \ldots, L-1$$

where $0<\alpha<1$ and $-1<\beta_1<\infty$ $(\beta_1 \neq 0)$ are parameters. We then assume that his utility function for nondurable purchases and durable holdings for his entire lifetime is

$$(6) \quad U_i(\cdot) = \kappa \sum_{j=i}^{L-1} (1+\gamma)^{j-i} u(n_j, d_j)^{-\beta_2}$$

where $\gamma > -1$, $-1 < \beta_2 < \infty$ $(\beta_2 \neq 0)$, and κ are parameters. The utility function (6) is a monotonic transformation of Kazuo Sato's two-level *CES* production function.[10] The restrictions on the parameters given above plus the restriction that $\kappa\beta_2 < 0$ assure that (6) is strictly quasi concave with positive marginal utilities of nondurable purchases and durable holdings. In (5) the parameter α indicates the distribution of utility between the two goods in any period. In (6) the parameter γ is the individual's subjective rate of time preference.

The interpretation of the parameters $\sigma_1 = 1/(1+\beta_1)$ and $\sigma_2 = 1/(1+\beta_2)$ is best obtained by following Sato's terminology. He calls σ_1 the "intraclass" elasticity of substitution. It is the elasticity of substitution within a particular class of goods, ignoring any substitutability among classes. Thus, for our model σ_1 can be interpreted as the elasticity of substitution of nondurable and durable services at any age when any intertemporal substitution effects are ignored. Sato calls σ_2 the "interclass" elasticity of substitution. It is the elasticity of substitution among input groups. Thus, σ_2 is the elasticity of substitution between

---

[9] This approach of deriving explicit functional forms using explicit utility functions has been used before. A list of references is given on pp. 591–92 of my earlier consumption function study.

[10] The same kind of utility function is used by Laurits Christensen in his analysis of the effects of the rate of return on saving. The inputs in his utility function are consumption and leisure.

$$(7a) \qquad k_j(\cdot) = \left[\frac{\alpha}{1-\alpha}\rho_j\right]^{-\sigma_1} \qquad j = i, \ldots, L-1$$

$$(7b) \qquad h_j(\cdot) = 1 + \rho_j k_j(\cdot) \qquad j = i, \ldots, L-1$$

$$(7c) \quad g_i(\cdot)^{-1} = h_i(\cdot)^{(\sigma_2-\sigma_1)/(1-\sigma_1)} \sum_{j=i}^{L-1} h_j(\cdot)^{(1-\sigma_2)/(1-\sigma_1)} \left\{\left[\left(\frac{1}{1+r}\right)(1+\gamma)^{1/\beta_2}\right]^{j-i} (p_j^n/p_i^n)\right\}^{1-\sigma_2}$$

where $$\rho_j = \left[p_j^d - p_{j+1}^d\left(\frac{1-\delta}{1+r}\right)\right] \Big/ p_j^n$$

pairs of either $(n_j, n_{j'})$, $(d_j, d_{j'})$, or $(n_j, d_{j'})$ where $j \neq j'$.[11]

We now define three additional variables to aid in the presentation of the explicit demand functions which result from the maximization of (6) subject to the intertemporal wealth constraint (2a′). These new variables are shown in (7a)–(7c). Using this notation, the individual's explicit purchase demand function for nondurables and his reservation demand function for durables in the current period are

$$(8a) \qquad n_i = g_i(\cdot)w_i/p_i^n$$

$$(8b) \qquad d_i = k_i(\cdot)n_i$$

His total dollar expenditures for durable and nondurable purchases are[12]

$$(9) \quad p_i^n n_i + p_i^d d_i = [1 + (p_i^d/p_i^n)k_i(\cdot)]g_i(\cdot)w_i$$

Thus, from (8) and (9) we see that $g_i(\cdot)$ is the individual's marginal propensity to purchase nondurables out of total current wealth, that $k_i(\cdot)$ is his desired ratio of durable holdings to nondurable holdings, and that $[1+(p_i^d/p_i^n)k_i(\cdot)]g_i(\cdot)$ is his marginal propensity to purchase out of current wealth.

B. *Expectations and Unobservable Variables*

A final step remains before the demand functions (8a) and (8b) can be estimated. These demand functions have future purchase prices and future labor income (through human wealth) as exogenous variables. Since these variables are not directly observable, we must relate them to observable variables in order to estimate the demand equations. Specific assumptions concerning the way in which the representative consumer forms his expectations are part of our maintained hypothesis.

We will very simply assume that the representative consumer expects some constant rate of inflation or deflation π^n in the purchase price of nondurables and some constant rate of inflation or deflation π^d in the purchase price of durables. Of course, these expected rates of price change do not need to be the same. Under these assumptions, the future purchase prices of nondurables and durables are related to current (observable) purchase prices according to

---

[11] The relationship between σ_1 and σ_2 and the Allen partial elasticities of substitution is discussed by Sato.

[12] This is not, however, the individual's nominal consumption during age i which is given instead by his total dollar expenditures on *services* of nondurables and durables during the period $(q_i^n n_i + q_i^d d_i)$. The difference between consumption and expenditures arises because the service price of durables is not equal to the purchase price of durables since the individual can purchase durables in order to rent them to himself in more than one period. His consumption of durables in any period is therefore equal to the amount of "rental income" which he pays himself in that period. This rental income does not bear any direct relation to his purchases of durables during that period. As Michael Farrell and Daniel Suits correctly point out, it is consumer expenditures and not consumption which affect employment, production, and income in the economy.

$$\text{(10a)}\quad p_j^n = (1+\pi^n)^{j-i} p_i^n \quad j = i+1, \ldots, L-1$$

$$\text{(10b)}\quad p_j^d = (1+\pi^d)^{j-i} p_i^d \quad j = i+1, \ldots, L-1$$

Further, we assume that the representative consumer forms his expectations concerning future nominal labor income in very much the same way that he forms his expectations concerning future prices. We assume that he expects his nominal future labor income to grow at the constant rate $\xi+\pi$ over the remainder of his economic lifetime. Under this assumption his future nominal labor income is related to his current nominal labor income according to

$$\text{(10c)}\qquad y_j = (1 + \xi + \pi)^{j-i} y_i \qquad j = i + 1, \ldots, L - 1$$

In this formulation ξ reflects expected changes in nominal labor income caused by changes in productivity and π reflects nominal labor income growth caused by expected secular inflation. When π is viewed in this light, it is obvious that this inflation rate need not be the same as the anticipated inflation rates for either nondurable or durable prices.

The assumptions (10a)–(10c), however, do not completely solve our unobservable variables problem, but only serve to reduce the number of unobservable variables in the model since the three expected inflation rates are themselves not observable. We must now consider the problem of how the representative consumer forms his expectations concerning these rates of inflation. One possibility which has been used a great deal in the literature is to assume that the consumer uses an adaptive expectations mechanism to form his expectations. Suppose that we are considering the consumer's expectations concerning the rate of inflation in the price on nondurables. Suraj Gupta has shown that the adaptive expectations framework can be written as

$$\text{(11)}\qquad \hat{\pi}^n = \phi[ln\,(p_i^n) - ln\,(\hat{p}_i^n)]$$

where $\hat{p}_i^n$ is the price of nondurables which the consumer expected to hold at age i. This expectations framework states that if the current price level is above the "normal" price level which the consumer expected to prevail in this period, then he will expect the price level to continue to increase in the future. Gupta goes on to show that (11) can be written as

$$\text{(12)}\qquad \hat{\pi}^n = (1 - \lambda) \sum_{j=0}^{T} \lambda^j \pi_{i-j}^n, \qquad T \leq i$$

where $\lambda = e^{-\phi}$ and π_{i-j}^n is the inflation rate which the individual observed at age $i-j$. Thus, under the adaptive expectations scheme the individual forms his expectation of the rate of inflation which will occur in the future by taking a weighted average of all past inflation rates.

However, Gupta proposes a second expectations scheme, namely,

$$\text{(13)}\qquad \bar{\pi}^n = \phi'[ln\,(\hat{p}_{i+1}^n) - ln\,(p_i^n)]$$

This expectations framework states that if the current price level is above the normal price level which the consumer expects to prevail in the future, then the consumer expects a negative rate of inflation as the price level returns to its normal level.

There is, of course, no a priori way to choose between these two hypotheses. However, Gupta is able to show that

$$\text{(14)}\qquad \bar{\pi}^n = \hat{\pi}^n[\phi'(\phi - 1)/\phi]$$

indicating that these two expectations hypotheses will be negatively related if $0<\phi<1$, the generally expected case.

Therefore, in order to allow for either of these cases to occur we will assume that the consumer forms his expectations concerning the rate of inflation of the price of nondurables according to

$$(15a) \qquad 1 + \pi^n = \theta_1 + \theta_2\hat{\pi}^n$$

This formulation has two advantages. First of all, it allows either (11) or (13) to obtain through the parameter θ_2. If $\theta_2 > 0$, then this formulation argues that the consumer is forming his expectations according to (11). However, if $\theta_2 < 0$, then it argues that the consumer is forming his expectations according to (13). Secondly, it allows us to test whether changes in inflation rates have an effect on consumer expenditures by testing whether or not θ_2 is significantly different from zero.

The constant θ_1 in (15a) is included to allow for the fact that the consumer may have an expectation of a nonzero long-term normal rate of inflation which is not influenced by any changes in past inflation rates. In other words, if the individual has a normal inflation rate which is not affected by current or past inflation rates, then $\theta_2 = 0$. However, in this case $(1+\pi^n)$ does not necessarily have to equal unity. Including θ_1 in the analysis takes this eventuality into account. We expect $\theta_1 > 0$, since it is improbable that the consumer would have a long-run expectation of a deflation rate of 100 percent per year.

We will use exactly the same procedure to relate the consumer's expectations concerning the rate of inflation in the price of durables and the rate of inflation which he sees as relevant to his nominal labor income growth. Consequently, we include the assumptions

$$(15b) \qquad 1 + \pi^d = \theta_3 + \theta_4\hat{\pi}^d$$

$$(15c) \qquad 1 + \xi + \pi = \theta_5 + \theta_6\hat{\pi}$$

as a part of the maintained hypothesis of our model. We expect θ_3 and θ_5 to be strictly positive.

Next, we consider the nominal interest rate to be used in the empirical analysis. The representative consumer may have long-run expectations concerning nominal interest rates which should be included in his demand functions. However, these expectations may or may not be sensitive to changes in current market interest rates, and they may not be solely related to the current level of an arbitrary market interest rate. In addition, they may be formed either adaptively in a manner similar to (11) or according to Gupta's alternative expectations scheme in a manner similar to (13). Therefore, to take account of these possibilities we redefine the nominal interest rate used in empirical analysis along the lines used for the inflation rates. Consequently, we include the assumption

$$(15d) \qquad 1 + r = \theta_7 + \theta_8\hat{r}$$

$$\text{where } \hat{r} = \left[(1-\lambda')\sum_{j=0}^{T} (\lambda')^j r^m_{i-j}\right], \qquad T \leq i$$

and r^m_{i-j} is a particular market interest rate which the individual observed at age $i-j$, as a part of our maintained hypothesis. This formulation affords the same advantages with respect to the nominal interest rate as assumptions (15a)–(15c) do with respect to the inflation rates. The constant θ_7 allows the consumer to have a nonzero long-run expected interest rate even if his expectation is not affected by changes in current market interest rates. The constant θ_8 allows us both to test whether interest rate expectations are influenced by changes in current market interest rates and to discriminate between the competing expectations hypotheses (11) and (13).

There is one final data problem which we must consider. The data available on the prices of nondurables and durables are price index data. These data cannot be used directly to measure the relative price of nondurables and durables since the ratio of the price indices of the two commodities is not necessarily equal to the relative price of the two commodities. Consequently, we assume that relative prices in the model are related to the ratio of the price indices according to

$$\frac{p_i^d}{p_i^n} = \theta_9 + \theta_{10} P^I \tag{15e}$$

where P^I is the ratio of the price index of durables to the price index of nondurables. The formulation (15e) allows the effects of relative price changes on the consumer's demands for nondurables and durables to be statistically tested through θ_{10}.

The problem of not being able to obtain data on relative prices also affects the model in a second respect. If we measure units of nondurables and durables purchased by dividing dollar expenditures by the price indices for the two commodities, we run into the problem that the constant dollar expenditures thus obtained are not necessarily in correct units. To circumvent this problem we will assume that the price index for nondurables is the "true price" of nondurables and then express the demand for durables in *units of nondurables* by dividing dollar demands for durables by the price index for nondurables.

If we substitute (10) and (15) into (8), we would obtain the individual's demand functions for nondurables and durables in terms of only observable variables. However, in the resultant demand functions we would encounter an "identification" problem. This problem is that in these demand functions the effects of one or more of the parameters in the model cannot be distinguished empirically. Consequently, it is necessary to reparameterize the model as follows. Define the vector of parameters $\underline{z} = (z_1, \ldots, z_{10})$ to be:

$$\begin{aligned}
z_1 &= \theta_7/[\theta_3(1-\delta)] \\
z_2 &= \theta_8/[\theta_3(1-\delta)] \\
z_3 &= \theta_4/\theta_3 \\
z_4 &= \theta_9\left(\frac{\alpha}{1-\alpha}\right)^{-\sigma_1/(1-\sigma_1)} \\
z_5 &= \theta_{10}\left(\frac{\alpha}{1-\alpha}\right)^{-\sigma_1/(1-\sigma_1)} \\
z_6 &= \theta_1/\theta_3 \\
z_7 &= \theta_2/\theta_3 \\
z_8 &= \theta_7/[\theta_1(1+\gamma)^{1/\beta_2}] \\
z_9 &= \theta_6/\theta_5 \\
z_{10} &= \theta_7/\theta_5
\end{aligned} \tag{16}$$

In terms of this new notation, the explicit demand functions for nondurables and durables of the individual in terms of only observable variables are

$$n_i = g_i^e(\cdot) w_i^e / p_i^n \tag{8a$'$}$$

$$d_i^* = k_i^e(\cdot) n_i \tag{8b$'$}$$

where $d_i^* = p_i^d d_i / p_i^n$ and $g_i^e(\cdot)$, $k_i^e(\cdot)$, and w_i^e denote $g_i(\cdot)$, $p_i^d k_i(\cdot)/p_i^n$, and w_i after substituting (15) and (16).

The demand functions (8a′) and (8b′) will form the basis of our empirical analysis.[18] These demand functions have the property that both the level and composition of the consumer's expenditures may be affected by changes in nominal interest rates, by changes in the rate of inflation in the price of durables, and by changes in the current relative prices of durables and

[18] For the most part, the demand for purchase of consumer durables has been estimated using a stock adjustment model similar to that of R. Stone and D. A. Rowe. See, for example, the studies by Michael Hamburger, David Huang, and De-Min Wu. This model states that consumers purchase durables in order to reduce the gap between their current holdings of durables and their desired stock of durables by some percentage. Some of the reasons advanced for incomplete adjustment are habit persistence and financial considerations. Our formulation of the model in terms of the intertemporal utility maximization problem of the individual has taken habit persistence and financial considerations into account and eliminated the need to use the stock adjustment framework. Thus, the individual's demand to purchase durables in our model is the difference between his reservation demand given by (8b′) and his current holdings of durables.

nondurables. The level of his expenditures, but not their composition, may be affected by changes in the rate of inflation of the price of nondurables.

III. An Empirical Test of the Model

The model of consumer demand for nondurables and durables is estimated using annual aggregate data for the United States for the period 1930–70, omitting 1941–46, for a total of 35 observations. The observation period ends in 1970, the last year for which wealth data were available. The aggregate demand for purchase of nondurables and aggregate reservation demand for durables equations which are used as the basis of the empirical estimation are obtained by assuming that (8a′) and (8b′) are the aggregate demand functions when the appropriate planning horizon is assumed.[14] This aggregation assumption, of course, assumes that there are no distribution effects and no demographic effects which would affect aggregate demand. The planning horizon which is used in the analysis is 30 years, which is approximately the life expectancy of an individual whose age is equal to the median age of household heads over the observation period.

Purchases of nondurables are defined to be consumer purchases of nondurables and services. The reservation demand for durables is the sum of current purchases of durables plus beginning of period durable holdings.[15] The beginning of period household wealth variable is the sum of their holdings of land, residential structures, consumer durables, and net financial assets plus the after-tax nonlabor income of both proprietors and nonproprietors. The latter is included to take account of the interest earned on the beginning of period wealth, which is paid during the current period. After-tax labor income is the sum of the after-tax labor income of proprietors and nonproprietors.

Price indices for nondurables and durables are used in the calculation of the relative price variable and the inflation rates of nondurables prices and durables prices. The inflation rate used in the calculation of expected labor income growth is based upon the *GNP* deflator. Two market interest rates are used in the empirical analysis. They are the Moody's Baa corporate bond yield and the Durand's basic yield on corporate bonds with five years to maturity.[16]

The series for the expected inflation rates $\hat{\pi}^n$, $\hat{\pi}^d$, and $\hat{\pi}$ and the expected interest rate series $\hat{r}$ were all calculated from (12) according to

$$(17) \qquad x_t^e = (1-\lambda)x_t + \lambda x_{t-1}^e \quad t = 1930\text{–}70$$

where x_t and x_t^e are the actual and expected inflation or interest rate series in period t, respectively. For any given x_t series and initial value x_0^e for x_{1929}^e, the choice of λ will determine a unique series for x_t^e. All of the expected inflation and interest rate series used in this study were constructed under the assumption that $x_0^e = x_{1930}$, the actual

---

[14] An alternative aggregation procedure would be to sum the demand functions (8a′) and (8b′) over the various age groups in the economy. However, estimating the resulting aggregate demand functions would require data on nonhuman wealth and labor income by age group. Such data are not available.

[15] Consumer rentals of the services of durables from the business sector are included in the purchases of services. However, the amounts of these rentals as measured by the sum of "Laundering in Establishments," "Taxicabs," and one-third of "Automobile Repair, Greasing, Washing, Parking, Storage and Rental" (Table 2.5, *The National Income and Product Accounts of the United States, 1929–1965*) are both a small and decreasing share of total expenditures on services and are small relative to purchases of durables. Therefore, the inclusion of these rentals in our data on nondurable purchases should not affect our empirical results. Further, purchases of durables for the purpose of renting them to others should be included in the national income statistics as purchases of producer durables. Thus, the potential double-counting problem mentioned in fn. 5 does not arise with these data.

[16] A more complete discussion of the sources of the data and the construction of the data will be given on request.

TABLE 1—MAXIMUM LIKELIHOOD ESTIMATES OF (8a′) AND (8b′)
$\lambda=0.50$

	Interest Rate Series			
Parameter	Durand's Basic Yield		Moody's Baa	
σ_1	0.801	(0.987)	0.716	(0.836)
		(0.049)		(0.078)
σ_2	0.750	(0.816)	0.562	(0.744)
		(0.406)		(0.402)
z_1	47.2	(150)	13.5	(39.5)
		(8.61)		(6.16)
z_2	−7.42	(0.022)	−3.25	(−1.02)
		(−25.8)		(−9.79)
z_3	0.157	(0.523)	0.170	(0.381)
		(−0.233)		(−0.057)
z_4	1.01	(1.44)	1.13	(1.54)
		(0.893)		(0.741)
z_5	−0.751	(−0.346)	−0.797	(−0.442)
		(−0.822)		(−1.06)
z_6	0.664	(1.22)	0.866	(1.08)
		(0.643)		(0.751)
z_7	1.03	(2.00)	0.340	(1.66)
		(−0.248)		(−0.344)
z_8	1.83	(3.46)	1.30	(4.01)
		(1.34)		(1.15)
z_9	0.070	(0.238)	0.041	(0.206)
		(−0.091)		(−0.097)
z_{10}	1.08	(1.11)	1.07	(1.11)
		(1.04)		(1.02)
$T^{-2}\|\hat{\Omega}\|$	0.105×10^4		909	
$T^{-1}\Sigma e_{1t}^2$	28.1		22.4	
$T^{-1}\Sigma e_{2t}^2$	117		130	
$T^{-1}\Sigma e_{1t}e_{2t}$	−47.4		−44.9	
$\Sigma e_{1t}^2/\Sigma(n_t-\bar{n})^2$	0.004		0.003	
$\Sigma e_{2t}^2/\Sigma(d_t-\bar{d})^2$	0.017		0.019	

inflation rate or interest rate which prevailed in 1930.[17] The x_t^e series used in the study were also generated for the entire time period 1930–70 to avoid any discontinuities due to the omission of the 1941–46 observations in the empirical analysis. Expected inflation rate and interest rate series generated using several different values of λ were tried in the empirical analysis to determine which value of λ yielded the best fit.

Maximum likelihood estimates of the model (8a′) and (8b′) are presented in Table 1 for both nominal interest rate series.[18] All of the expected interest rate and inflation series used in the estimation were calculated with $\lambda = 0.5$. Other values of λ were tried in other estimations, but the resultant sums of squared residuals were larger than that obtained using the value of 0.5 for λ. The numbers in parentheses to the right of the parameter estimates are approximate upper and lower bounds for 99 percent likelihood ratio confidence intervals on the estimates.[19]

---

[17] Of course, the choice of the initial value of x_0^e is arbitrary and does affect the actual expected series generated. M. Feldstein and G. Chamberlain encounter the same problem and assume that $x_0^e=0$. They, however, have 24 periods of experience preceding the first observation in their sample. I chose my initial condition as described above because I did not have the good fortune of Feldstein and Chamberlain. The reader is referred to their discussion of the general estimation problems involved in using this approach and the references contained therein.

[18] Maximum likelihood estimates of the parameters of the model were obtained as follows. Let $\underline{\epsilon}_t=(e_{1t}, e_{2t})'$ be the vector of error terms in period t and add e_{1t} and e_{2t} to (8a′) and (8b′), respectively. Assume that the $\underline{\epsilon}_t$ are identically and independently distributed bivariate normal random variables with mean vector $\underline{0}$ and covariance matrix $\underline{\Omega}$. Then under this assumption the maximum likelihood estimates of the parameters are obtained by minimizing: (over)

The results indicate that the demand functions fit the data quite well. For the nondurables demand function (8a′), the mean squared errors are 28.1 and 22.4 for the Durand's basic yield and the Moody's Baa rate, respectively. For the durable holdings equation (8b′), the corresponding mean squared errors are 117 and 130. These are quite low when compared to the deviation of the dependent variables about their means. These comparisons are given in the last two rows of the table. These results show that the fit with the nondurable expenditure equation is better than that with the durable holdings equation. Further, the results in the row labelled $T^{-1}\Sigma e_{1t}e_{2t}$ indicate that on the average the errors in the expenditure equations are such that if the regression overpredicts nondurable purchases it will underpredict durable holdings and vice versa.

The results in Table 1 indicate that both the intraclass and interclass elasticities of substitution (σ_1 and σ_2, respectively) are less than unity and the confidence intervals on the estimates indicate that we can reject the null hypotheses that they are equal to zero or one at the 0.01 level of significance. The results on σ_1 indicate that for the consumer sector durable holdings and nondurable purchases in any time period are poor substitutes when intertemporal substitution effects are ignored. The results on σ_2 indicate that for the consumer sector nondurables are poor intertemporal substitutes, durable holdings are poor intertemporal substitutes, and durable holdings in one period are poor substitutes for nondurable purchases in another period.

Because of the sign restrictions on α, δ, θ_3, and θ_5, we know from (16) that the signs of θ_2, θ_4, θ_6, and θ_8 will be the same as the signs of z_7, z_3, z_9, and z_2, respectively. Thus, the results in Table 1 indicate that θ_2, θ_4, and θ_6 are all positive and that θ_8 is negative. The results on θ_2, θ_4, and θ_6 imply that the consumer sector forms its expectations concerning inflation rates according to the adaptive expectations framework (11). The result on θ_8 implies that the consumer sector has an expectation of a long-run normal nominal interest rate, and that if the current nominal interest rate rises above this level, consumers expect interest rates to fall in the future.

These results appear contradictory because increases in the rate of inflation should, in the long run, cause increases in the nominal rate of interest if the real rate of interest remains unchanged. However, the confidence intervals on the parameters z_3, z_7, and z_9 include zero which indicates that we cannot reject at the 0.01 significance level the null hypothesis that con-

---

$$|\Omega| = \left(\sum_{t=1}^{T} e_{1t}^2\right)\left(\sum_{t=1}^{T} e_{2t}^2\right) - \left(\sum_{t=1}^{T} e_{1t}e_{2t}\right)^2$$

where e_{jt} is the residual of equation j in time period t, and T is the number of observations. See E. Malinvaud, pp. 568–70.

[19] The use of likelihood ratio confidence intervals is not standard in the literature. However, a Monte Carlo study by Robert Blattberg, Timothy McGuire, and the author found that the estimated small sample power functions for the likelihood ratio test always tend to unity. Power functions for hypothesis tests based upon estimated variances using the inverse of the information matrix or the inverse of the matrix of second partials of the sum of squared residuals did not. These results suggest that confidence intervals based upon the matrix inverses may be too large in some instances. Likelihood ratio confidence intervals also need not be symmetric as must confidence intervals based upon matrix inverses. Heuristically, there seems to be no justification for restricting confidence intervals to be symmetric in non-linear models. A more extensive treatment of the likelihood function and shortest average confidence intervals is given in S. Wilks.

sumers have constant, long-run expectations concerning inflation rates which are unaffected by changes in current inflation rates. If this is the case, then there is no contradiction at all. This type of expectation concerning inflation rates is consistent with the type of expectations behavior found concerning nominal interest rates.[20]

The results of the empirical investigation to determine whether changes in nominal interest rates, relative purchase prices, and inflation rates affect consumer expenditures are presented in Table 2.[21] Likelihood ratio test statistics are given for various null hypotheses concerning the statistical significance of these variables as determinants of consumer expenditures. These hypothesis tests are suggested by the following results of the theoretical discussion of the model: nominal interest rates will be a determinant of expenditures only if $z_2 \neq 0$. Relative prices will be a determinant of expenditures only if $z_5 \neq 0$. The rates of inflation of nondurables and durables will be determinants of expenditures only if $z_7 \neq 0$ and $z_3 \neq 0$, respectively. The *GNP* deflator will affect expenditures only if $z_9 \neq 0$.

In the first row of Table 2, we test the null hypothesis that the model explains consumer expenditures as well leaving out interest rate, relative price, and inflation rate effects as it does including them. The very high value of the test statistics under this hypothesis indicates that we can reject this null hypothesis at better than the 0.001 level of significance for both interest rate series.

TABLE 2—LIKELIHOOD RATIO STATISTICS FOR TESTS OF SIGNIFICANCE OF NOMINAL INTEREST RATES, RELATIVE PRICES, AND INFLATION RATES

Null Hypothesis	Durand's Basic Yield	Moody's Baa
$z_2 = z_3 = z_5 = z_7 = z_9 = 0$ $(\hat{r}; P^I; \hat{\pi}^d; \hat{\pi}^n; \hat{\pi})$	38.0[a]	43.1[a]
$z_2 = 0$ $(\hat{r})$	6.52[b]	11.3[a]
$z_5 = 0$ (P^I)	24.5[a]	25.5[a]
$z_3 = 0$ $(\hat{\pi}^d)$	1.15	2.81[c]
$z_7 = 0$ $(\hat{\pi}^n)$	3.54[c]	1.78
$z_9 = 0$ $(\hat{\pi})$	0.593	0.094
$z_3 = z_7 = 0$ $(\hat{\pi}^d; \hat{\pi}^n)$	4.10	3.64

[a] Significant at the 0.001 level.
[b] Significant at the 0.025 level.
[c] Significant at the 0.1 level.

In the second row of the table we test the null hypothesis that nominal interest rates are not a statistically significant determinant of expenditures. We find that we can reject this null hypothesis at the 0.001 level when the interest rate series in the empirical analysis is Moody's Baa corporate bond yield and that we can reject this hypothesis at the 0.025 level when the interest rate series is the Durand's basic yield. The null hypothesis that relative prices are not a statistically significant determinant of consumer expenditures is tested in row 3 of the table. For both interest rate series this null hypothesis is rejected at better than the 0.001 level.

The null hypothesis that the inflation rates of nondurable prices, durable prices, and the *GNP* deflator are not determinants of consumer expenditures are tested in rows 4, 5, and 6 of Table 2, respectively. None of these hypotheses can be rejected at even the 0.05 level of significance. Because there was one case in which either the rate of inflation of durable prices or the rate of inflation of nondurable prices was significant at the 0.1 level, we decided to test the further null hypothesis that the inflation rates of durable prices and nondurable prices are simultaneously not determinants of expenditures. The result of this hypothesis test is given in the last row of the table and we find that we cannot reject this null hypothesis at even the 0.1 level of significance.

---

[20] There is a second possible explanation for these results. Robert Lucas and Leonard Rapping state that "... there is little evidence that nominal interest rates adjust to expected inflation with sufficient speed to maintain a constant real rate" (p. 276n). If this is the case, then consumers may not have had enough experience with increases in inflation rates increasing nominal interest rates in the period covered by our observations for this relationship to be part of their expectations structure. The reader is referred to the references in Lucas and Rapping to empirical studies yielding the evidence.

[21] Throughout this study we have referred to (8b′) as a demand for durable holdings equation. However, if the stock of durables held coming into the period is constant, then an increase in the demand to hold durables must give rise to a demand to purchase durables. Hence, our use of this equation to discuss expenditure changes.

Thus, the overall results of the hypothesis tests concerning the significance of nominal interest rates, relative prices, and inflation rates as determinants of consumer expenditures indicate that changes in nominal interest rates and relative prices do influence consumer expenditures, but that changes in inflation rates do not.

The effects of changes in relative prices, interest rates, and inflation rates on the model are given in Table 3. Here it is found that an increase in the relative price of durables P^I has the expected effect of increasing the marginal propensity to purchase nondurables and to lower the ratio of desired durable holdings to nondurable purchases. The results of these effects are seen to be that consumers increase the expenditures on nondurables and decrease their durable purchases. However, the overall effect is to increase total expenditures.

The results in Table 3 indicate that an increase in the weighted average of current and past nominal interest rates ($\hat{r}$) has the effect of decreasing the marginal propensity to purchase nondurables and slightly increasing the desired ratio of durable holdings to nondurable purchases. Further, the results indicate that this increase in $\hat{r}$ increases consumers' expected total human and nonhuman wealth (w^e). The overall effect of an increase in $\hat{r}$ is to increase consumer expenditures on durables and nondurables.

TABLE 3—RELATIVE PRICE, NOMINAL INTEREST RATE, AND INFLATION RATE EFFECTS (BOTH INTEREST RATE SERIES)

	Δg^e	Δk^e	Δw^e	Δn	Δd	$\Delta(n+d)$
ΔP^I	+	−	0	+	−	+
$\Delta \hat{r}$	−	+[a]	+	+	+	+
$\Delta \hat{\pi}^d$	−	+	0	−	−	−
$\Delta \hat{\pi}^n$	−	0	0	−	−	−
$\Delta \hat{\pi}$	0	0	+	+	+	+

[a] Positive, but very small.

In order to explain these results it is necessary to recall the type of expectations behavior concerning interest rates indicated by the parameter estimates in Table 1. The results indicated that consumers view an increase in the weighted average of current and past nominal interest rates as indicating that nominal interest rates will be lower in the future. These expected lower interest rates increase the present value of the stream of future labor income and cause consumers to increase their expenditures due to the higher value of their human wealth.

An alternative explanation of this interest rate effect is that a high current nominal interest rate which raises $\hat{r}$ indicates that in the current period consumers can earn a higher than their expected normal rate of return on bonds. Consequently, they can achieve the same real income in the future with a smaller amount of bond purchases. They respond to this opportunity by increasing the expenditures on nondurables and durables in the current period.

The results of the hypothesis tests in Table 2 indicate that changes in current inflation rates are not statistically significant determinants of expenditures. However, we also present the signs of the effects of changes in these rates indicated by the parameter estimates given in Table 1 as a check on our empirical results. Since the results indicate that consumers form expectations concerning the rate of inflation adaptively, we would expect signs of changes in $\hat{\pi}^n$ and $\hat{\pi}^d$ on consumer expenditures to be the opposite of those of $\hat{r}$. We would also expect that an increase in $\hat{\pi}^d$ would cause consumers to increase their desired ratio of durable holdings to nondurable purchases because it lowers the price of durables *services* relative to nondurables *services*. Further, increases in $\hat{\pi}$ indicate that consumers expect future secular increases in their nominal labor income which we would expect to cause them to increase their current period expenditures. The results in Table 3 are consistent with all of these a priori expectations.

IV. Summary and Conclusions

In this paper we constructed an intertemporal utility maximizing model of consumer behavior for the case in which consumers can purchase both durable and nondurable goods. From this model we were able to obtain a reservation demand function for current durable holdings and a demand function for current nondurable purchases which included total human and nonhuman wealth, nominal interest rates, relative prices, and inflation rates as independent variables. Further, the model was constructed in such a way that it was possible to test whether nominal interest rates, relative prices, and inflation rates were statistically significant determinants of consumer expenditures.

An empirical estimation of the model over the period 1930–70, omitting 1941–46, and using aggregate data for the United States indicated that nominal interest rates and relative prices were statistically significant determinants of consumer expenditures. However, the empirical results also indicated that the null hypothesis that changes in inflation rates do not affect expenditures cannot be rejected.

The results of the empirical estimation were also used to determine the direction of nominal interest rate and relative price changes upon expenditures. Here it was found that an increase in the weighted average of current and past nominal interest rates would increase consumer expenditures on both nondurables and durables. This finding is in accord with that of my earlier study of consumption. It was also found that an increase in the relative price of durables would cause consumers to decrease their expenditures on durables and to increase their expenditures on nondurables. The overall effect of the increase would be to increase total expenditure.

REFERENCES

A. Ando and F. Modigliani, "The 'Life-Cycle' Hypothesis of Saving: Aggregate Implications and Tests," *Amer. Econ. Rev.*, Mar. 1963, *53*, 55–84.

R. Blattberg, T. McGuire, and W. Weber, "A Monte Carlo Investigation of the Power of Alternative Techniques for Testing Hypotheses in Nonlinear Models," mimeo., 1971.

L. Christensen, "Saving and the Rate of Return," unpublished doctoral dissertation, Univ. California 1968.

F. de Leeuw and E. Gramlich, "The Federal Reserve-MIT Model," *Fed. Reserve Bull.*, Jan. 1968, *54*, 11–40.

W. E. Diewert, "Intertemporal Consumer Theory and the Demand for Durables," *Econometrica*, May 1974, *42*, 497–516.

M. Farrell, "The New Theories of the Consumption Function," *Econ. J.*, Dec. 1959, *69*, 678–96

M. Feldstein and G. Chamberlain, "Multimarket Expectations and the Rate of Interest," *J. Money, Credit, Banking*, Nov. 1973, *5*, 873–902.

I. Fisher, *The Rate of Interest*, New York 1907.

M. Friedman, *A Theory of the Consumption Function*, Princeton 1957.

S. Gupta, "The Portfolio Balance Theory of the Expected Rate of Change of Prices," *Rev. Econ. Stud.*, Apr. 1970, *37*, 187–203.

M. Hamburger, "Interest Rates and Demand for Consumer Durables," *Amer. Econ. Rev.*, Dec. 1967, *57*, 1131–53.

D. Huang, "Discrete Stock Adjustment: The Case of Demand for Automobiles," *Int. Econ. Rev.*, Jan. 1964, *5*, 46–62.

R. Lucas and L. Rapping, "Real Wages, Employment and Inflation," *J. Polit. Econ.*, 1969, *77*, 721–54; reprinted in E. Phelps, ed., *Microfoundations of Employment and Deflation Theory*, New York 1970, 257–305.

E. Malinvaud, *Statistical Methods of Econometrics*, Chicago 1966.

F. Modigliani and R. Brumberg, "Utility Analysis and Aggregate Consumption Functions: An Attempt at Integration," mimeo., 1953.

——— and ———, "Utility Analysis and the Consumption Function: An Interpretation of Cross Section Data," in K. Kurihara, ed., *Post-Keynesian Economics*, New Brunswick 1954.

R. W. Parks, "The Demand and Supply of Durable Goods and Durability," *Amer. Econ. Rev.*, Mar. 1974, *64*, 37–55.

K. Sato, "A Two-Level Constant-Elasticity-of-Substitution Production Function," *Rev. Econ. Stud.*, Apr. 1968, *34*, 201–18.

R. Stone and D. A. Rowe, "The Market Demand for Durable Goods," *Econometrica*, July 1957, *25*, 423–43.

D. B. Suits, "The Determinants of Consumer Expenditure: A Review of Present Knowledge," in *Impacts of Monetary Policy: Commission on Money and Credit*, Englewood Cliffs 1963.

D. B. Suits and G. Sparks, "Consumption Regressions with Quarterly Data," in J. S. Duesenberry et al., eds., *The Brookings Quarterly Model of the United States*, Chicago 1965.

W. Weber, "The Effect of Interest Rates on Aggregate Consumption," *Amer. Econ. Rev.*, Sept. 1970, *60*, 591–600.

S. Wilks, "Shortest Average Confidence Intervals from Large Samples," in *Annals Mathemat. Statist.*, 1938, *9*, 166–74..

D-M. Wu, "An Empirical Analysis of Household Durable Goods Expenditure," *Econometrica*, Oct. 1965, *33*, 761–80.

U.S. Office of Business Economics, *The National Income and Product Accounts of the United States, 1929–1965*, Washington 1966.

Exchange Rates and Fiscal Policy in a Popular Model of International Trade

By RUDIGER DORNBUSCH*

This paper develops a reinterpretation of the devaluation analysis in a popular model of international trade developed by Charles Bickerdike, Joan Robinson, and Lloyd Metzler. It has come to be known as the "elasticity approach."[1] That model, developed in terms of independent markets for imports and exports, has been criticized for its apparent lack of general equilibrium properties. The assumptions underlying the model have remained by and large implicit and vary across users, and the integration of this analysis with the "absorption approach" and macroeconomics has never been demonstrated. All this is very much recognized and indeed application of the model is consistently accompanied by appropriate "partial equilibrium," "impact effect," or "upper bound" caveats. These shortcomings notwithstanding, the model continues to enjoy substantial popularity in policy discussions and interpretations of current events, in empirical work, and in recent textbooks in the field of international economics.[2]

The *BRM* model is likely to remain the preferred tool in the analysis of trade balance issues—along with the foreign trade multiplier model. It is therefore interesting and useful to enquire into its formal general equilibrium implications. This paper is concerned with one such interpretation involving a nontraded goods sector and a well-defined government policy that makes the analysis consistent with the absorption approach. While it may well be possible to generate alternative interpretations, the present analysis would seem to have the advantage of simplicity while at the same time pointing out the essential properties that would have to be satisfied by alternative treatments.

In Section I a brief review of the *BRM* approach is offered. In Section II that approach is interpreted in terms of a simple model with nontraded goods and constant terms of trade. The implicit assumptions of the *BRM* model are brought to the foreground by an explicit consideration of the market for home goods, the equilibrium conditions, and the role of the relative price of nontraded goods.[3] Following that analysis, Section III considers the subsidiary issues that are raised by variable terms of trade. The notion of "total" elasticities is given an interpretation in Section IV where we return to the composite traded good model to explore alternative assumptions about absorption policies.

I. The Bickerdike-Robinson-Metzler Model

The *BRM* model discusses the effects of exchange rate changes in terms of separate markets for imports and exports. In each market, demand and supply depend only

---

* Associate professor, Massachusetts Institute of Technology. I am indebted to Ronald Jones and Michael Mussa for helpful discussions.

[1] In this analysis it will be referred to as the *BRM* model.

[2] See, for example, the articles by William Branson and Stephen Magee (1973, 1975); and the textbooks by Robert Heller, Mordechai Kreinin, and Robert Stern.

[3] Throughout this paper the terms home goods and nontraded goods are used interchangeably. This use differs from the Keynesian tradition where home goods are identified with exportables because the constancy of absolute and hence relative prices of domestically produced goods allows them to be aggregated.

on the *nominal* price in terms of the importing or exporting country's currency, and cross-price effects between markets do not exist. The model is defined in equations (1) to (5):

$$M(P_m) = X^*(P_m^*) \tag{1}$$

$$X(P_x) = M^*(P_x^*) \tag{2}$$

$$B = P_x X - P_m M \tag{3}$$

$$P_m = P_m^* e \tag{4}$$

$$P_x = P_x^* e \tag{5}$$

where[4]

M, X = domestic imports and exports
M^*, X^* = foreign imports and exports
P_m, P_x = the domestic currency prices of (domestic) imports and exports
P_m^*, P_x^* = the foreign currency prices of imports and exports
e = the domestic currency price of foreign exchange

Equation (1) is the equilibrium condition in the market for the home country's imports, and equation (2) is the equilibrium condition for the export market. Equation (3) defines the home country's trade balance surplus measured in terms of domestic currency. Equations (4) and (5) relate the domestic and foreign currency prices of goods via the exchange rate.

The initial equilibrium is shown in Figure 1 at the intersection of the solid demand and supply schedules drawn for an exchange rate e^0. At that initial equilibrium the value of imports $P_m^0 M^0$ is equal to the value of exports $P_x^0 X^0$ so that the trade balance is in equilibrium.

Consider now a devaluation by the home country. The home demand for imports and supply of exports, drawn against their domestic currency prices, will remain unaffected. The foreign supply and demand schedules, however, will shift. In order to maintain the foreign currency price of goods, by (4) and (5) the domestic currency price would have to rise in the same proportion as the exchange rate. It follows that both the foreign supply and demand schedule shift upward in the same proportion as the rate of devaluation. The new equilibrium obtains where both markets clear. The domestic price of imports and exports increases to P_m^1 and P_x^1, respectively; imports decline to M^1 and exports expand to X^1.

The effect of a devaluation on the trade balance in this model is ambiguous. While the value of exports increases, the value of imports may rise or fall depending on the elasticity of demand. A sufficient condition for trade balance improvement is provided by the familiar *BRM* condition.[5]

The disturbing features of this model arise primarily from its apparent disagreement with general equilibrium analysis. The particular questions that arise are the following:

> Why should the behavioral equations depend on a single money price, rather than on relative prices and real income or real expenditure.
>
> The budget constraint would imply that not all markets can be independent so that there must either be additional markets not accounted for explicitly in the model or else the model violates the budget constraint.
>
> It is well known from the absorption approach that a trade balance surplus corresponds to an excess of income over expenditure; to the extent that a devaluation in this model generates a trade balance surplus it must at the same time lower expenditure relative to income in a manner that is in no way explained in equations (1)–(5).

---

[4] All starred variables or quantities denote the foreign country.

[5] See equation (20′) below.

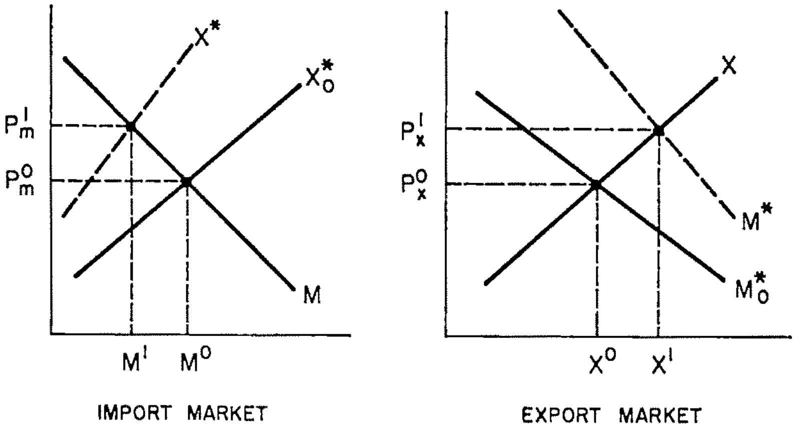


Figure 1

None of these questions is particularly novel; indeed it is well understood by now that there are some things in the background and that some of these things are being held constant. Since these background stories are essential to the operation of the model, it is important to establish what exactly is implicitly assumed.

One interpretation takes a monetary view of the problem and assumes that there is another good in each country, money, and that the nominal quantity of money is fixed. The real effects of a devaluation derive in this formulation from the effect on expenditure of the reduction in the real value of money attendant upon a devaluation.[6]

An alternative and more accepted interpretation assumes that there is another commodity in each country, a nontraded good or home good, and that there is an active policy of maintaining constant some nominal quantity. Among the nominal quantities to be held constant, the following have been alternatively suggested: money income; money expenditure; wages; the price level; or the prices of a particular class of commodities.[7] It is this latter interpretation that we want to develop by an explicit consideration of the market for nontraded goods. That analysis will show that the formulation in Figure 1 can be interpreted to assume the following:[8]

> A nominal price of home goods that is maintained constant in each country by fiscal policy;[9]
>
> A unit marginal propensity to spend on nontraded goods;[10]

---

[6] For such an interpretation see Murray Kemp (1969), Takashi Negishi, and the author (1973).

[7] See in particular the articles by Charles Kennedy, Sidney Alexander, Fritz Machlup, June Flanders, Gottfried Haberler, Ronald Jones (1961), and the book by James Meade.

[8] Kemp (1962, 1969) provides an identical list.

[9] The discussion in this paper is developed in terms of fiscal policy although what is essentially required is a policy that will control absorption so that alternatively the argument could be developed in terms of monetary policy. In the present context we assume (implicitly) that a Meade-type "neutral monetary policy" is pursued so as to maintain constant the rate of interest and thereby allow us to disregard the interrelationship between goods and asset markets. For a further discussion, see the author (1974).

[10] An alternative view that interprets the behavioral equations as reduced forms and thereby allows for positive marginal propensities to spend on traded goods is developed in Section IV. The assumption of a zero marginal propensity to spend on traded goods is obviously quite extraordinary and the reader is reminded that I am not constructing the preferred model of the trade balance, but rather seek to explain the *BRM* model.

Zero cross-price effects (substitution effects) between traded goods.

This set of assumptions, as will be shown below, ensures that there is no feedback effect from the "background" and that fiscal policy eliminates any market disequilibrium that arises in the nontraded goods market. It explains at the same time that the trade balance effects of a devaluation are due to the fiscal policy a devaluation brings into operation.

II. A Simple Model with Nontraded Goods

In this section a simple model is developed with nontraded goods. It is assumed that the relative price of traded goods is constant so that there is only one composite traded good in the world and one nontraded good in each country. The assumption of one composite traded good is a convenience that allows us to disregard terms of trade issues and instead focus consideration on the relative price of traded goods in terms of home goods and on absorption.[11]

Figure 2 shows the Salter diagram with the full employment transformation curve VV between traded goods production Q_I and nontraded goods Q_N, as well as indifference curves. The shape of the indifference curves reflects the assumption that the marginal propensity to spend on home goods is unity so that the indifference curves are laterally parallel. Full equilibrium obtains at point E where income equals expenditure and where the demand and supply of traded and nontraded goods are respectively equal so that both the home-goods market clears and the trade balance is in equilibrium.

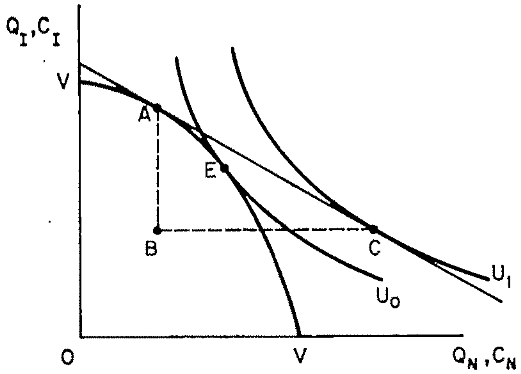


FIGURE 2

An increase in the relative price of traded goods moves the production equilibrium from E to A while moving consumption from E to C; productive resources move into the traded goods industry while consumers substitute from traded to nontraded goods. At that higher relative price there is an excess demand for nontraded goods equal to BC, and an excess supply of traded goods equal to AB. We note, too, that the trade balance surplus AB, measured in terms of nontraded goods, is equal to the excess demand for nontraded goods. The latter equality is an implication of the assumption that income equals expenditure.

Consider now the operation of fiscal policy. We will assume that fiscal policy is used to maintain fixed the *nominal* price of home goods. This implies that any changes in the nominal prices of traded goods, as brought about by an exchange rate change, constitute changes in the relative prices of traded goods in terms of home goods. It implies, too, that the government has to levy income taxes whenever there is an excess demand for home goods; the increase in income taxes reduces private expenditure thereby eliminating the excess demand and preventing the price of home goods from rising. Conversely, when there is an excess supply of nontraded goods so that their price would tend to decline, income transfers are required so as to raise private expenditure and eliminate the excess supply.

The assumption that the marginal propensity to spend on home goods is equal to one becomes important now, since it implies that the tax collection in response to an excess demand has to be identically equal to the excess demand. This can be seen from Figure 2 by noting that in order to eliminate the excess demand for nontraded goods BC, consumption has to move to point B, which requires that the budget constraint facing consumers pass through that point, which in turn implies that disposable income declines by the distance BC due to the income taxes.

---

[11] The model developed here is a two-country version of W. E. G. Salter's "dependent economy" model.

A further implication of the unit marginal propensity to spend on home goods is the fact that the fiscal policy will have no effect on the trade balance. The trade balance surplus will depend only on the nominal price of traded goods.

We conclude that we can represent the home country's real trade balance surplus $\overline{B}$ as an increasing function of the nominal price of traded goods P:

$$\overline{B} = \overline{B}(P) \tag{6}$$

By an identical line of argument the foreign country's real trade balance surplus $\overline{B}^*$ is an increasing function of the foreign nominal price of traded goods P^*.

$$\overline{B}^* = \overline{B}^*(P^*) \tag{7}$$

Finally the nominal prices of traded goods are related by the exchange rate e:

$$P = P^*e \tag{8}$$

Given the policy of pegging the nominal prices of nontraded goods in both countries, an exchange rate change becomes a change in a real variable. Devaluation of the home currency raises the *relative* price of traded goods in terms of nontraded goods at home and lowers the *relative* price of traded in terms of nontraded goods abroad. In response to that change in relative prices, resources move into the traded goods sector at home while consumers substitute from traded goods to home goods. Abroad exactly the opposite move takes place: productive resources move to the home-goods sector while consumers substitute toward traded goods thus generating a trade balance surplus for the home country. The magnitude of the surplus and the relative price changes will depend on the substitutability between goods.

The equilibrium condition in the world traded goods market is that one country's excess supply of traded goods equal the other country's excess demand for traded goods. Using (6), (7), and (8) this equilibrium condition may be written as follows:

$$\overline{B}(P) = -\overline{B}^*(P/e) \tag{9}$$

Figure 3 shows the determination of the equilibrium in the traded goods market. The home country's excess supply of traded goods or trade balance surplus $\overline{B}(P)$ is drawn as an increasing function of the domestic currency price of traded goods. Similarly foreign excess demand for traded goods is shown as a decreasing function of the domestic price of traded goods for a given exchange rate e^0. Initial equilibrium obtains at a domestic price of traded goods P^0 where the world traded goods market is in equilibrium and trade is balanced. Returning to Figure 2, this implies that income equals expenditure in

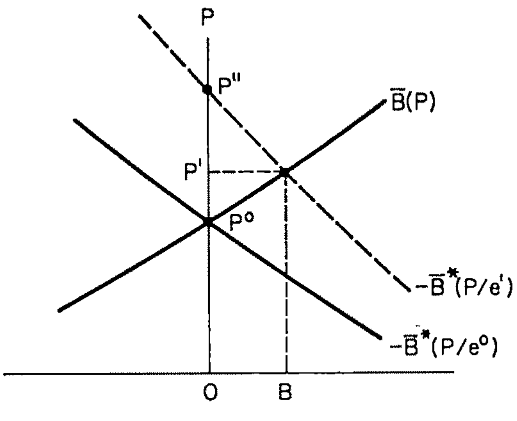


FIGURE 3

each country and that the home-goods market clears in both countries.

A devaluation by the home country will change the relation between domestic and foreign prices. In terms of Figure 3, the foreign excess demand schedule shifts upward in the proportion of the rate of devaluation. At the initial domestic currency price of traded goods P^0 there is a world excess demand for traded goods since the foreign price of traded goods has declined in the proportion as the exchange rate. In response to that excess demand the price of traded goods increases until the world market clears at a new domestic currency price P'. At that new price the home country has a trade balance surplus OB equal to the foreign country's deficit.

We note that the increase in the domestic price of traded goods $(P'-P^0)/P^0$ falls short of the rate of devaluation $(P''-P^0)/P^0$ which implies that the foreign price of traded goods has declined in the proportion $(P''-P')/P''$. It follows that a devaluation raises the domestic (relative) price of traded goods and lowers the foreign (relative) price of traded goods.

To appreciate why a devaluation exerts real effects, it is obviously necessary to go to the background of Figure 3 and study the fiscal policies that are triggered off by these price changes. In the home country, as was shown in Figure 2, the increase in the relative price of traded goods causes an excess demand for home goods which in turn is eliminated by an income tax collection equal to AB when measured in terms of traded goods. This implies that the counterpart of the trade balance surplus is an identically equal tax collection or budget surplus. It follows that the analysis is fully consistent with the absorption approach since it readily explains the decline in expenditure relative to income implied by the trade balance surplus. It follows, too, that in the absence of this fiscal policy the price of home goods would increase in response to the excess demand until relative prices returned to their initial position with no real effects persisting.

Consider next what happens in the foreign country. Abroad the nominal price of traded goods declines, and given the nominal price of home goods, the relative price of traded goods declines. In response to that relative price change there will be an excess supply of nontraded goods triggering off income transfers in order to restore equilibrium in the nontraded goods market. Figure 4 shows the new equilibrium in the foreign country with production at point A, consumption at point B, and a trade balance deficit equal to AB. The excess of expenditure over income or the income transfer measured in terms of traded goods is equal to the trade balance deficit.

Having shown that the domestic trade balance surplus is equal to income tax collections and that abroad the deficit equals the income transfers made by the government, it remains to infer from the equality of one country's deficit and the other country's surplus that a devaluation operates in the same manner as an international transfer from one country to another. Indeed, in this model a devaluation has exactly the same effects as a policy of transferring AB units of real income from the home country to the foreign country.

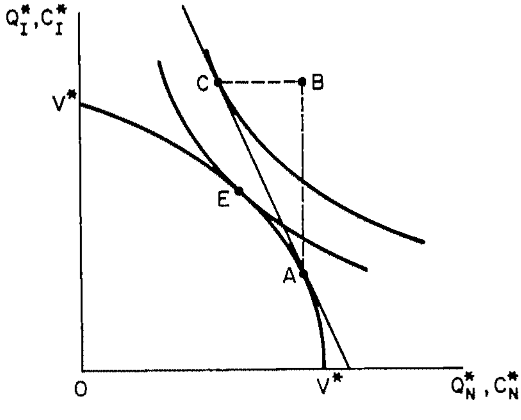


FIGURE 4

The preceding discussion can be formalized with the help of a few definitions. The real trade balance measured in terms of exportables is equal to the supply of exports less the demand for imports measured in terms of exportables:

$$(6') \quad \bar{B}(P) \equiv X(P_x) - (P_m/P_x)M(P_m)$$

where we assume that P_m/P_x is constant so that by an appropriate choice of units we have $P_m = P_x = P$. Similarly the foreign trade balance surplus $\bar{B}^*$ is defined by

$$(7') \quad \bar{B}^*(P^*) \equiv (P_m^*/P_x^*)X^*(P_m^*) - M^*(P_x^*)$$

Next we define the *compensated* elasticities of demand for imports and supply of exports:

$$(10) \quad \hat{M} = -\eta\hat{P}_m; \qquad \hat{X} = \epsilon\hat{P}_x;$$

$$\hat{M}^* = -\eta^*\hat{P}_x^*; \qquad \hat{X}^* = \epsilon^*\hat{P}_m^*$$

where a ˆ denotes a proportional change in a variable.

Using these definitions, assuming initial trade balance equilibrium and constant terms of trade, we can define the elasticity of the trade balance surplus or the excess supply of traded goods with respect to the relative price of traded goods in terms of home goods:

$$(11) \quad \frac{d\bar{B}}{dP}\,\frac{P}{X} = \sigma \equiv \epsilon + \eta;$$

$$\frac{d\bar{B}^*}{dP^*}\,\frac{P^*}{M^*} = \sigma^* \equiv \epsilon^* + \eta^*$$

We can now differentiate the equilibrium condition in (9), and using the definitions in (11) we derive the effect of a devaluation on the domestic (relative) price of traded goods:

$$(12) \quad \hat{P} = \frac{\sigma^*}{\sigma + \sigma^*}\,\hat{e}$$

Differentiating equation (8) and substituting from (9) yields the effect of a devaluation on the foreign price:

$$(13) \quad \hat{P}^* = \frac{-\sigma}{\sigma + \sigma^*}\,\hat{e}$$

We note from (12) and (13) that the distribution of price changes between countries will depend on the substitutability between traded goods and home goods in the two countries.

To determine the effect of the devaluation on the home country's trade balance we differentiate (6'), and using (11) and (12) we have:

$$(14) \quad d\bar{B} = X\,\frac{\sigma\sigma^*}{\sigma + \sigma^*}\,\hat{e}$$

Equation (14) shows that a devaluation unambiguously improves the trade balance of the devaluing country, and thus confirms the results derived in Figure 3.

The fact that a devaluation cannot worsen the trade balance distinguishes this composite traded good version most fundamentally from the more general form of the *BRM* model. It points out that the possibility of a worsening of the trade balance is not due to a devaluation per se, but rather to the terms of trade changes that may accompany a devaluation. The model developed here emphasizes, too, that as long as the terms of trade remain unchanged the scope for "elasticity pessimism" is very limited. We note from (14) that at worst a devaluation will leave the trade balance unchanged, a circumstance that arises if in either country there is zero substitutability between home goods and traded goods.

In concluding this section we remark that the essential elements of the *BRM* model are contained in the present version that maintains constant the terms of trade. These essential elements are that a devaluation changes the relative price of traded goods in terms of home goods thereby generating a surplus in the devaluing country and a deficit abroad. The counterpart of the change in the trade balance arises in the home-goods markets: an excess demand at home and an excess supply abroad. It is essential to the success of a devaluation that there be an "internal balance" policy that validates the relative price changes by a reduction in absorption at home and an increase in absorption abroad.

The fact that terms of trade changes are "almost" secondary in their role in this context has been recognized by I. F. Pearce (1961) and June Flanders who emphasize both the requirement of absorption changes and the change in the relative price of traded goods in terms of home goods. Regrettably, the terms of trade may change, and therefore a more complicated analysis is called for.

III. The Terms of Trade, Transfers, and the Trade Balance

The preceding analysis showed that at constant terms of trade a devaluation has the same effects as an income transfer: it lowers domestic absorption and increases foreign absorption where the magnitude of the absorption change is identically equal to the trade balance surplus shown in (14). In deriving the equilibrium change in absorption we have assumed that the world market for traded goods clears; we have not, however, shown that the market for each individual good will clear.

To determine whether the analysis in Section II is in fact consistent with equilibrium in each individual traded-goods market, it is sufficient to consider equilibrium in one of these markets. This is so since the preceding analysis called for clearing of the world traded-goods market so that by implication the excess demand in any one market would be offset by the excess supply in the other market. Without loss of generality we can conduct the analysis in terms of the market for the home country's importable. We will use the results of Section II to compute the change in domestic import demand and foreign export supply and determine whether they are exactly equal so that no change in the terms of trade is called for.

The change in the equilibrium price of importables, at constant terms of trade, is given by (12). Using the definition of the elasticity of demand for imports in (10) and substituting for the price change from (12) we have:

$$\hat{M} = -\eta\hat{P}_m = -\eta\frac{\sigma^*}{\sigma+\sigma^*}\hat{e} \tag{15}$$

The change in the foreign supply of exports is similarly derived by using the definition of the elasticity of supply in (10) and the equilibrium price change in (13), noting that $\hat{P}^*_m \equiv \hat{P}^*$.

$$\hat{X}^* = \epsilon^*\hat{P}^*_m = -\epsilon^*\frac{\sigma}{\sigma+\sigma^*}\hat{e} \tag{16}$$

The world excess demand for importables, at constant terms of trade, is given by the difference between (15) and (16), which with some manipulation may be written as:

$$dM - dX^* = X\frac{\sigma\sigma^*}{\sigma+\sigma^*}\left[\frac{\epsilon^*}{\epsilon^*+\eta^*} - \frac{\eta}{\epsilon+\eta}\right]\hat{e} \tag{17}$$

The home country's terms of trade will worsen or improve as there is at constant terms of trade a world excess demand or supply for importables, that is depending on the familiar Robinson criterion:

$$\epsilon\epsilon^* - \eta\eta^* \gtrless 0 \tag{18}$$

Equation (17) has a ready interpretation in terms of the transfer problem. The term $X\sigma\sigma^*/(\sigma+\sigma^*)$ measures the change in absorption or the change in the trade balance that is induced by a devaluation at constant terms of trade. In the home country the trade balance surplus is achieved through a reduction in imports and an increase in exports. The fraction of the reduction in absorption or the improvement in the trade balance that is reflected in reduced imports is $\eta/(\epsilon+\eta)$. Abroad the increased absorption and worsened trade balance is reflected in part in a decreased level of exports. The term $\epsilon^*/(\epsilon^*+\eta^*)$ measures the fraction of the increased absorption that is reflected in decreased exports. These fractions are not marginal propensities to spend that we ordinarily encounter in the discussion of the transfer problem, but they do play exactly the same role in that they indicate the effect of a change in absorption on individual commodity excess demands.[12] The channel here is not a direct expenditure effect measured by a marginal propensity, but rather a substitution effect that derives from the change in the equilibrium relative prices of home goods occasioned by the change in absorption that at constant relative prices falls entirely on these goods.

The preceding discussion shows that there is no presumption whatsoever about the direction in which a devaluation, or a transfer for that matter, will change the terms of trade. This point deserves emphasis since there continues to be the belief that a devaluation in and of itself changes the terms of trade. If a devaluation does, however, change the terms of trade the effects on the trade balance are no longer described by (14), but have to be derived from the disaggregated model in (1) to (5). Differentiating equations (1), (2), (4), and (5) and using the definitions in (10) yields the effects of a devaluation on the domestic prices of importables and exportables:

$$\hat{p}_m = \frac{\epsilon^*}{\epsilon^* + \eta}\hat{e}; \qquad \hat{p}_x = \frac{\eta^*}{\epsilon + \eta^*}\hat{e} \tag{19}$$

Differentiating the trade balance equation in (3) and substituting the equilibrium price changes from (19) gives the trade balance effect of a devaluation:[13]

$$dB = P_xX\left[(1+\epsilon)\frac{\eta^*}{\epsilon+\eta^*} - (1-\eta)\frac{\epsilon^*}{\epsilon^*+\eta}\right]\hat{e} \tag{20}$$

which with some manipulation may be rewritten in the more familiar *BRM* form:

$$dB = P_xX\left[\frac{\eta\eta^*(1+\epsilon+\epsilon^*) - \epsilon\epsilon^*(1-\eta-\eta^*)}{(\epsilon+\eta^*)(\epsilon^*+\eta)}\right]\hat{e} \tag{20'}$$

Inspection of (20) shows that if the terms of trade remain unchanged or improve, $\eta^*/(\epsilon+\eta^*) \geq \epsilon^*/(\epsilon^*+\eta)$, a devaluation will improve the trade balance. A worsening of the terms of trade, however, brings about the possibility of a worsening of the trade balance.[14] The interpretation of that possibility is as follows: The increase in the relative price of both traded goods in terms of home goods shown in (19) generates a substitution effect that induces a trade surplus and an excess demand for home goods. Against this substitution effect operates the income effect of a terms of trade deterioration which generates an excess supply of nontraded goods and a trade deficit. Similarly abroad the decline in the relative price of traded goods generates an excess supply of home goods on account of the substitution effect while the income effect of the foreign terms of trade improvement generates an excess demand. Whether the trade balance improves or worsens will depend on the relative size of the income and substitution effects. A more than sufficient condition is that in either country, the compensated elasticity of demand for imports be larger than unity —a condition that ensures that substitution effects dominate income effects.

---

[12] For a discussion of the transfer problem in a more general model with nontraded goods, see Ian McDougall, Pearce (1961, 1970), and Jones (1974).

[13] The reader will note that for the constant terms of trade case equation (20) reduces to (14) by using the substitution $\eta\eta^* = \epsilon\epsilon^*$.

[14] Jones (1974) has developed the interpretation of the *BRM* condition in terms of the stability of the underlying commodity markets.

The above argument shows, too, why a devaluation cannot worsen the trade balance when it improves the terms of trade. In this event both substitution effects and income effects operate in the same direction generating an excess demand for nontraded goods and as its mirror image a trade surplus; abroad, similarly, both the income and substitution effects operate in the same direction to generate an excess supply of nontraded goods and a corresponding trade deficit. Furthermore, if the terms of trade remain unchanged there are only substitution effects and hence an unambiguous improvement in the trade balance.

The possibility that a devaluation worsens the trade balance cannot be ruled out on theoretical grounds. Before taking alarm at that contingency, however, it is worth noting what the implications for the nontraded goods markets are. If that should not be sufficient to discourage "elasticity pessimism," it should still be borne in mind that the model contains some assumptions that may be inordinately strong, in particular, the unit marginal propensity to spend on home goods and the zero cross elasticity between traded goods.

IV. An Alternative Interpretation

In his discussion of the elasticity approach, Alexander pointed out that the import and export equations of the *BRM* model could be viewed as reduced forms that embody the effects of the absorption policies pursued in the background and where these effects are not necessarily limited to substitution effects as we have assumed so far. On this alternative interpretation the elasticities are no longer compensated elasticities but rather "total" elasticities, and it becomes important to know what effects they embody. One rather straightforward interpretation suggested by the analysis of Section II is developed here. On the assumption that the terms of trade remain constant we can interpret the response to a change in the price of traded goods as composed of both a substitution effect and an expenditure effect. The expenditure effect derives now from the assumption that the marginal propensity to spend on traded goods is positive so that the fiscal policies pursued in the background exert a direct effect on the trade balance.

The trade balance surplus can now be written as a function of the (relative) price of traded goods and disposable income $Y-T$, where Y and T denote respectively the value of output and taxes both measured in terms of home goods:[15]

$$\tilde{B} = \tilde{B}(P, Y - T) \tag{21}$$

Differentiating (21) and using (11) we have:[16]

$$d\tilde{B} = \sigma X\hat{P} + (1 - \pi)dT/P \tag{22}$$

where $(1-\pi)$ denotes the marginal propensity to spend on traded goods.

As was shown in Section II the trade surplus is equal to the change in absorption which in turn is equal to the change in taxes. We can therefore equate in (22) the terms $d\tilde{B}=dT/P$ to obtain the reduced form elasticity of the trade balance with respect to the relative price of traded goods:[17]

---

[15] The value of output measured in terms of home goods is defined as $Y=PQ_I+Q_N$.

[16] A change in the relative price of home goods has no income effect since demand equals production. Accordingly (22) only reflects substitution effects and absorption effects of a tax change.

[17] The equality of the change in the trade balance and in taxes is derived from the budget constraint. The value of expenditure equals the value of output less taxes: $PC_I+C_N=PQ_I+Q_N-T$ or $P(Q_I-C_I)+(Q_N-C_N)=T$. When the home-goods market clears, $Q_N=C_N$, the trade surplus is equal to taxes or $\tilde{B}=(Q_I-C_I)=T/P$.

$$d\tilde{B}/X = (\sigma/\pi)\hat{P} \equiv \bar{\sigma}\hat{P} \tag{23}$$

We note that (23) differs from (11) by the term in the denominator and reduces to (11) if the marginal propensity to spend on home goods, π, is unity. The direct expenditure effects on traded goods exerted by the fiscal policy raises the trade balance effects of a relative price change.

Proceeding in a similar manner for the foreign country we derive the reduced form elasticity $\bar{\sigma}^* \equiv \sigma/\pi^*$. Substituting the terms $\bar{\sigma}$ and $\bar{\sigma}^*$ in (14) instead of the terms σ and σ^* yields the trade balance effect of a devaluation with positive marginal propensities to spend on traded goods:

$$d\tilde{B} = X\frac{\sigma\sigma^*}{\sigma\pi^* + \sigma^*\pi}\hat{e} \tag{24}$$

Equation (24) shows that the trade balance effects of a devaluation are larger the smaller the marginal propensities to spend on home goods in each country. The reason is simply that if less of a tax change falls on home goods then larger tax changes are required in order to maintain home-goods market equilibrium when the relative price of home goods is changed by an exchange rate change. Equation (24) therefore serves to highlight the role of the absorption policies in the *BRM* model.

So far the analysis has assumed that fiscal policy is used in each country to maintain constant the nominal price of home goods. This is certainly not the only interpretation of total elasticities. An alternative that has been widely referred to is the assumption that nominal income is maintained constant. This would require that an increase in the nominal price of traded goods be accompanied by a decline in the nominal prices of home goods. Assuming that fiscal policy maintains constant the nominal value of income, we have the following relationship between price changes:[18]

$$\hat{P}_N = -[(1-\theta)/\theta]\hat{P} \tag{25}$$

where $\hat{P}_N$ denotes the proportional change in nominal home-goods prices, and where θ denotes the share of home goods in nominal income.

The trade balance effect of a change in the nominal price of home goods is now composed of three elements: The substitution effects induced by a rise in traded goods prices (relative to home-goods prices) and the accompanying decline in home-goods prices and the effect of the fiscal policy required to maintain home-goods market equilibrium at a constant nominal income. We can therefore rewrite the trade balance effect of a change in the *nominal* price of traded goods as:

$$d\tilde{B}/X = \bar{\sigma}(\hat{P} - \hat{P}_N) = (\sigma/\pi\theta)\hat{P} \tag{26}$$

where we have substituted for the home-goods price change from (25).

Equation (26) shows that the trade balance effect of an increase in the nominal price of traded goods is inversely proportional to the share of home goods in nominal income. This is so because by (25) the smaller the share of home goods, the larger the change in the relative price of traded goods induced by a change in their nominal price, $\hat{P} - \hat{P}_N = (1/\theta)\hat{P}$. We can derive an expression analogous to (26) for the foreign country and substitute the reduced form or total elasticities, $\lambda \equiv \sigma/\pi\theta$ and $\lambda^* \equiv \sigma^*/\pi^*\theta^*$, in (14) to find that the effects of a devaluation on the trade balance are the larger the smaller the shares of home goods in nominal income, given a policy in each country of maintaining constant money income:

$$d\tilde{B} = X\frac{\sigma\sigma^*}{\sigma\pi^*\theta^* + \sigma^*\pi\theta}\hat{e} \tag{27}$$

---

[18] From the definition of nominal income $Y = PQ_I + P_NQ_N$, we derive equation (25) by differentiation. A policy that maintains constant nominal income is equivalent to maintaining constant a price index $\bar{P} = P^{(1-\theta)}P_N^{\theta}$. For a further discussion, see the author (1974) and Pearce (1970). Since the nominal price of home goods now is flexible, P only denotes the nominal price of traded goods.

A comparison of equations (24) and (27) shows that the real effects of a devaluation depend crucially on what nominal variables are held constant. Thus a policy of maintaining constant the nominal prices of home goods generates a smaller trade surplus than one that maintains constant nominal income. Since exchange rate changes exert real effects because they constitute changes relative to some other nominal variable, it is indispensable to ask what that variable is in order to determine the nature and magnitude of the response to an exchange rate change.

V. Concluding Remarks

This paper has developed an interpretation of the *BRM* model that gives emphasis to the general equilibrium context in which the analysis of the traded-goods markets has to be placed. Such a general equilibrium model will involve a market for home goods and an explicit policy of maintaining constant some nominal variable. That policy provides the link with the absorption approach. Comparison of alternative policies shows that the real effects of an exchange rate change depend critically on the choice of nominal variables to be pegged. It points out, too, that there is not much use in considering reduced form equations without making explicit the general equilibrium and macroeconomic system from which they are derived.

REFERENCES

S. Alexander, "The Effects of Devaluation: A Simplified Synthesis of Elasticities and Absorption Approaches," *Amer. Econ. Rev.*, Mar. 1959, *49*, 21–42.

C. Bickerdike, "The Instability of Foreign Exchange," *Econ. J.*, Mar. 1920, *30*, 118–22.

W. Branson, "The Trade Effects of the 1971 Currency Realignments," *Brookings Papers*, Washington 1972, *2*, 15–69.

R. Dornbusch, "Currency Depreciation, Hoarding, and Relative Prices," *J. Polit. Econ.*, July/Aug. 1973, *81*, 893–915.

———, "Alternative Price Stabilization Policies and the Effects of Exchange Rate Changes," unpublished paper, Univ. Chicago, 1974.

J. Flanders, "The Balance of Payments Adjustment Mechanism: Some Problems in Model Building," *Kyklos*, 1963, *16*, 395–412.

G. Haberler, "The Market for Foreign Exchange and the Stability of the Balance of Payments," *Kyklos*, 1949, *3*, 93–218; reprinted in R. Cooper, ed., *International Finance*, Baltimore 1969.

R. Heller, *International Monetary Economics*, Englewood Cliffs 1974.

R. W. Jones, "Stability Conditions in International Trade: A General Equilibrium Analysis," *Int. Econ. Rev.*, May 1961, *2*, 199–209.

———, "Trade with Nontraded Goods: The Anatomy of Interconnected Markets," *Economica*, May 1974, *41*, 121–38.

M. Kemp, "The Rate of Exchange, the Terms of Trade and the Balance of Payments in Fully Employed Economies," *Int. Econ. Rev.*, Sept. 1962, *3*, 314–27.

———, *The Pure Theory of International Trade*, Englewood Cliffs 1969.

C. Kennedy, "Devaluation and the Terms of Trade," *Rev. Econ. Stud.*, 1950, no. 1, *18*, 28–41.

M. Kreinin, *International Economics*, New York 1971.

F. Machlup, "Relative Prices and Aggregate Expenditure in the Analysis of Devaluation," *Amer. Econ. Rev.*, June 1955, *45*, 255–79.

S. Magee, "Currency Contracts, Pass-Through and Devaluation," *Brookings Papers*, Washington 1973, *3*, 303–23.

———, "Prices, Incomes and Foreign Trade," in P. B. Kenen, ed., *International Trade and Finance: Frontiers for Research*, Cambridge 1975.

I. A. McDougall, "Non-Traded Goods and the Pure Theory of International Trade," in I. A. McDougall and R. N. Snape, eds., *Studies in International Economics*, Amsterdam 1970.

J. E. Meade, *The Theory of International Policy*, Vol. I: *The Balance of Payments*, Oxford 1951.

L. A. Metzler, "The Theory of International Trade," in H. Ellis, ed., *A Survey of Contemporary Economics*, Philadelphia 1949, 210–14; reprinted in his *Collected Papers*, Cambridge, Mass. 1973, 1–49.

T. Negishi, "Approaches to the Analysis of Devaluation," *Int. Econ. Rev.*, June 1968, *9*, 218–27.

I. Pearce, "The Problem of the Balance of Payments," *Int. Econ. Rev.*, Jan. 1961, *2*, 1–28.

———, *International Trade*, New York 1970.

J. Robinson, "The Foreign Exchanges," in H. Ellis and L. A. Metzler, eds., *Readings in the Theory of International Trade*, Homewood 1950, 83–103.

W. Salter, "Internal and External Balance: The Role of Price and Expenditure Effects," *Econ. Rec.*, Aug. 1959, *35*, 226–38.

R. W. Stern, *The Balance of Payments*, Chicago 1973.

Equities, Capital Gains, and the Role of Finance in Accumulation

By B. J. Moore*

The purpose of the present paper is to develop some implications of the introduction of corporate equities and capital gain income into macro-economic growth models. Capital gains received by wealth owners are shown to be related to the volume of corporate retentions and to the valuation ratio, the latter defined as the ratio of the market value of ownership claims on property income to the market value of real assets.[1]

Retained earnings and capital gains are a large and volatile component of a broadly defined measure of total personal income. It is surprising that most previous studies of household consumption and saving behavior have not dealt explicitly with either retained earnings or capital gains. A Keynesian approach, which includes disposable income and wealth as variables in the consumption function, distinguishes sharply between dividend income and retained earnings. In contrast a Fisherian approach, which describes consumption as the allocation of an individual's wealth over his lifetime, treats all property income the same way, and implies that the allocation of profits between retained earnings and dividends has no effect on aggregate saving. Both views represent the consequences of specifying consumer behavior without sufficiently considering the implications of the corporate form of business, and the distinction between labor and property income. The present paper argues that the propensity to consume out of capital gain income plays a critical role in the Keynesian-neoclassical controversy concerning the independence of investment from saving behavior.[2] The paper closes with some implications of the argument for the role of finance in the process of capitalist accumulation and for the theory of interest.

I. Equities and Capital Gains in a Macro-Economic Growth Model

In conventional growth models golden age defines a path in which output, capital stock, and effective labor (assuming constant returns to scale and Harrod-neutral technical progress) all grow at the same constant rate g. Such a balanced growth path represents a moving equilibrium configuration only if the wealth propensities of the society are satisfied as well as the saving, investment, and technological requirements. If k is the capital-output ratio, the share of output that must be saved and invested in new capital is gk. But it is not self-evident that the desired wealth-income ratios of wealth owners are such as to maintain the required capital-output ratio.

---

* Professor of economics, Wesleyan University. I would like to express my indebtedness to conversations with Paul Davidson, Mike Lovell, and Joan Robinson, none of whom is in any way inculpated by the ideas presented below. The paper was conceived while I was a Senior Visitor to the Economics Faculty at Cambridge, on sabbatical leave from Wesleyan.

[1] The use of the valuation ratio as a macro-economic concept was first introduced by Nicholas Kaldor (1966) in his "Neo-Pasinetti Theorem."

[2] For a presentation of modern growth theories which emphasizes the centrality of this distinction, see Jan Kregel. Recently Franco Modigliani has revised the *FMP* model to emphasize the role of consumption out of capital gains in the transmission mechanism of monetary policy. See also Paul David and John Scadding.

While the latter is conventionally stated in terms of saving, growth, and technological requirements, economic units in the aggregate must simultaneously be willing to own and accumulate the resulting capital stock.

In a one-asset world, physical capital is identical to private wealth and real investment is identical to saving. In golden age growth the desired wealth-income ratio is the same as the required capital-output ratio, and the desired saving-income ratio is identical to the required investment-output ratio. With the introduction of financial assets and government or foreign sectors, total private wealth may diverge from the capital stock, and private saving from capital formation. The government or the rest of the world may own a share of the economy's capital stock, or may issue "outside" financial assets that can satisfy private demand for wealth and saving. A number of authors have introduced money and bonds into growth models to illustrate how government financial policy can alter the golden age path that simultaneously satisfies technological requirements and thrift propensities.[3]

The introduction of corporate equities represents an alternative mechanism by which wealth-income and capital-output ratios can diverge. Consider a model in which all real capital K is owned by business corporations and financed solely by the issue of stock. Corporations are intermediate wealth entities, owned ultimately by households. Assume for simplicity that corporate equities are the only type of financial assets, so that total household wealth W is held in the form of equities. If E is the price of equities and N the number of equities in existence:

$$W \equiv EN \tag{1}$$

Total household wealth W bears some relationship v to the value of the real capital stock K owned by corporations $v=W/K$. The valuation ratio v represents the ratio of the market value of ownership claims to corporations (the present value of expected future income from share ownership) to the market value of corporate real assets:

$$W \equiv vK \tag{2}$$

Differentiating with respect to time

$$dW \equiv dN \cdot E + dE \cdot N \tag{3}$$

$$dW \equiv dK \cdot v + dv \cdot K \tag{4}$$

where d represents the net change in the variable over the time interval considered.[4]

Total net output and income in any period consists of labor income L received by households, and property income P received initially entirely by corporations. If s_B is the corporate retention ratio, retained earnings are s_BP, and dividends paid to household shareowners are $(1-s_B)P$. Corporate net investment, the change in the real capital stock $I=dK$, may be financed internally out of retained earnings s_BP or externally from new share issue $(E \cdot dN = iI)$ where i represents the proportion of net investment financed externally.

$$I \equiv s_BP + E \cdot dN \equiv s_BP + iI \equiv dK \tag{5}$$

From equations (3), (4), and (5) an expression may now be derived for capital gains G:

$$\begin{aligned} G &\equiv N \cdot dE \equiv dW - E \cdot dN \\ &\equiv s_BP + (v-1)dK + K \cdot dv \end{aligned} \tag{6}$$

Capital gains are equal to corporate retained earnings s_BP plus corporate net investment times the excess of the valuation ratio over unity $(v-1)dK$, plus or minus capital gains or losses due to changes in the valuation ratio in the period $K\cdot dv$. For the case in which the productivity of new investment is equal to the discount rate applied to future income from share ownership, v will equal one, so that corporate retentions will result in an accrual in share values equal to the capital gain income to stockholders.

---

[3] For a summary of this literature, see Jerome Stein.

[4] Equations (3) and (4) hold only as approximations and only when the changes are small. The exact expressions require the addition of the term $dN \cdot dE$ to equation (3) and $dK \cdot dv$ to equation (4). Stock splits are ignored so that N changes only with new issues. For a more extended analysis of the introduction of corporate equities and capital gains into macro-economic growth models, see the author (1973a).

Household comprehensive income, the net accretion of economic power over any period, is equal to labor income, dividends, and capital gains:

$$\begin{aligned}(7)\quad Y_H^c &\equiv L + (1 - s_B)P + G\\ &\equiv L + P + (v-1)dK + K\cdot dv\end{aligned}$$

Corporate retention policy thus determines only the form in which household income is received between dividends and capital gains. All income earned by corporations, whether paid out or retained, is simultaneously comprehensive income to household shareowners. Business retentions thus exert a potentially expansionary influence on aggregate demand, depending on household propensity to consume out of capital gain income. They represent income available to be spent twice, first as retained earnings of businesses which may be spent on investment goods, and then as capital gain income to household capitalists, which may be saved or spent on consumption goods.

Total comprehensive income of the household and business sector Y^c is equal to current output plus capital gains. For accounting consistency, comprehensive output must be redefined to include capital gains. Saving out of comprehensive income S^c is equal to investment plus capital gains. Saving represents a decision to accumulate wealth, and can be satisfied by real asset accumulation or by the revaluation of existing assets:

$$\begin{aligned}(8)\quad Y^c &= Y + G = L + P + G\\ &= C + I + G\\ (9)\quad S^c &= S_B^c + S_H^c = I + G\\ &= s_B P + vI + K\cdot dv\\ (10)\quad S_H^c &= vI + K\cdot dv = iI + G\end{aligned}$$

The rate of growth of share prices is equal to:

$$\begin{aligned}(11)\quad \frac{dE}{E} &= \frac{dK}{K} - \frac{dN}{N} + \frac{dv}{v}\\ &= \frac{dK}{K}\left(1 - \frac{i}{v}\right) + \frac{dv}{v}\end{aligned}$$

When the valuation ratio remains constant $(dv=0)$, this is identical to the rate of growth of dividends per share. If corporate investment is expected to grow at the constant rate g in perpetuity $(g=I/K = dK/K)$, and future dividend and capital gain income is discounted at the constant rate r, this leads to the valuation formula:

$$(12)\quad E = \frac{(1 - s_B)P/N}{r - g\left(1 - \frac{i}{v}\right)}$$

Rewriting

$$(13)\quad r = \frac{(1 - s_B)P}{EN} + g\left(1 - \frac{i}{v}\right)$$

$$\begin{aligned}(13')\quad r &= \frac{P}{EN} + g\left(\frac{v-1}{v}\right)\\ &= \frac{1}{v}\left[\left(\frac{P}{K}\right) + g(v-1)\right]\end{aligned}$$

The discount rate and so the rate of return on equities is equal to the dividend yield plus the rate of growth of dividends per share. Alternatively the return on equities is equal to the earnings-price ratio plus the expression $(v-1)/v$ times the growth of corporate assets, which will be positive, zero, or negative depending on whether $v \gtreqless 1$. When $v=1$ the discount rate on shares is simply equal to the profit rate, and so the rate of return on investment.

II. Wealth-Income Ratios in Balanced Growth

The introduction of equities permits the value of private wealth to diverge from the value of the capital stock, so that saving diverges from real capital accumulation. If capital investment grows autonomously at the rate g $(I=gK)$, it is possible to formulate the relationship between private wealth and the real capital stock in balanced growth. A necessary condition for balanced growth is that the profit rate $(\pi=P/K)$, the rate at which the stock market discounts future income r, and the valuation ratio v remain constant. In this case $dv=0$, and capital gains and wealth accruals are equal to $G=s_BP+(v-1)gK$.

Assume first that households save some constant proportion (s_H) of their comprehensive income. The household saving function may be written:

$$(14)\quad \overset{o}{S_H} = s_H[L+(1-s_B)P+G] = s_H[Y+(v-1)I]$$

In balanced growth comprehensive saving and investment must be equal, so that it is possible to solve for the equilibrium value of v:

$$(15)\quad s_BP+s_H[Y+(v-1)I] = I+G = I+s_BP+(v-1)I$$

$$(15')\quad s_H[Y+(v-1)I] = vI$$

$$(16)\quad v = \left(\frac{s_H}{1-s_H}\right)\left(\frac{Y-gK}{gK}\right)$$

$$\frac{\partial v}{\partial s_H} > 0, \quad \frac{\partial v}{\partial g} < 0, \quad \frac{\partial v}{\partial\left(\frac{K}{Y}\right)} < 0$$

The valuation ratio, which expresses the wealth-capital ratio W/K, is positively related to household saving preferences, and negatively related to the growth rate and to the capital-output ratio. If investment is autonomous, and if household saving is independent of income form, corporate retention policy is irrelevant and has no effect on any real variables, but only determines the form in which household comprehensive income is received between dividends and capital gains. The valuation ratio will be greater or less than one depending on whether the household saving ratio out of comprehensive income is greater or less than the investment income ratio of business corporations ($v \gtrless 1$ according to whether $s_H \gtrless gK/Y$).

Household saving is in fact dependent on the form in which income is received.[5] Assume that the household saving function may be written:

$$(17)\quad \overset{o}{S_H} = s_LL+s_P(1-s_B)P+s_GG$$

where $1 \geq s_G \geq s_P \geq s_L \geq 0$.

The equilibrium value of v then becomes

$$(18)\quad v=\frac{s_LY+(s_P-s_L)P+(s_G-s_P)s_BP-s_GgK}{gK(1-s_G)}$$

$$\frac{\partial v}{\partial s_L}>0, \quad \frac{\partial v}{\partial s_P}>0, \quad \frac{\partial v}{\partial s_G}>0,$$

$$\frac{\partial v}{\partial s_B}\geq 0, \quad \frac{\partial v}{\partial g}<0, \quad \frac{\partial v}{\partial K}<0$$

The valuation ratio now depends positively on household and business saving propensities. The effect on v of corporate retentions depends on the size of the differential $(s_G - s_P)$, so that the effect on v of the share going to property depends on the differential $(s_P - s_L)$. New issues by firms will lower total saving and the valuation ratio so long as household saving out of capital gains exceeds saving out of dividend income $(s_G > s_P)$.

---

[5] Several empirical studies have estimated that the propensity to consume property income is less than the overall propensity to consume (see Lawrence Klein and Arthur Goldberger, Robert Holbrook and Frank Stafford). There have been few empirical studies of the effect of capital gains on consumption other than as a component of wealth, with a coefficient between 0.05 and 0.01 (see John Arena, Kul Bhatia, Franco Modigliani). In a recent paper Martin Feldstein has estimated that the propensity to consume the retained earnings component of capital gains was 0.25, while capital gains in general have a much smaller effect (0.02) on current consumption.

Empirical studies indicate that only a small proportion of capital gains and losses are realized (see Martin Bailey, Bhatia), and that the volume of unrealized gains has little impact on household consumption expenditures (see Arena). Alternatively expressed, the propensity to save out of capital gain income is extremely high $(s_G > .95)$. This suggests that changes in equity prices and in v must operate primarily through their effect on investment expenditures rather than consumption expenditures in order to restore equilibrium in the market for current output.

A. *The Significance of the Valuation Ratio*

Equations (16) and (18) state an extremely important relationship. Since households are the ultimate wealth owners under the institution of private property, the total value of the economy's wealth held in household portfolios must adjust until it satisfies the amount households desire to hold, which is in turn related to their (lifetime) income, the growth rate, and their saving preferences. So long as investment expenditures are autonomous and independent of financial forces, a change in household saving preferences by assumption cannot affect the rate of real capital formation. An autonomous rise in household saving propensities would result solely in a rise in the valuation ratio, and so in the ratio of comprehensive income to current output. So long as $s_G < 1$ this would raise household consumption expenditures as a proportion of current output, until equilibrium in the goods and security markets had been restored. As shown by equation (15′) v must adjust until household saving out of comprehensive income is equal to v times the rate of corporate real investment.

The introduction of equities and capital gains thus provides an additional escape from the Harrod-Domar knife-edge dilemma. In the simplest formulation $S = sY$ and $I = gK$, so that for full employment equilibrium investment and output must grow at the warranted rate $g_w = s/(K/Y)$. The neoclassical solution is to introduce a production function with variable factor proportions, so that the capital-output ratio K/Y adjusts over the long run to saver preferences to ensure that the warranted equals the natural rate. The neo-Keynesian solution is to introduce two sectors, capitalists and workers, with different saving propensities, so that the distribution of income between labor and property, the rate of profit, and the aggregate saving ratio(s) adjust to ensure equality of the warranted and natural rate (see Nicholas Kaldor, 1956). In the present capital gains model the valuation ratio, the ratio of wealth to income, the relation between comprehensive income and current output, and hence the proportion of current output consumed, adjust to satisfy full employment balanced growth. While the capital-output ratio may be technologically determined, the ratio of wealth to income, and hence the value of claims to the capital stock, is governed by saving preferences, since it must be consistent with the amount that wealth owners desire to hold. Saving represents the demand to accumulate assets, and these may be provided by real capital accumulation or by the revaluation of existing assets.

Changes in v thus introduce an additional stabilizing factor into the model, in the sense of an additional adjustment mechanism to equate planned saving by households and planned investment by corporations. If households desire to save more than corporations plan to invest, the result is to raise v and the value of wealth portfolios ($W=vK$), and thus provide the increase in wealth values that households desire. The deflationary implications of a rise in household savings are offset by a higher v and a higher ratio of comprehensive income and so consumption to current output. Conversely a decision by households to save less and accumulate wealth at a lower rate need not necessitate a lower rate of accumulation of real capital, but merely a reduction in v. An excess of planned investment over planned saving causes v to fall to a lower level, reducing household comprehensive income and so offsetting the expansionary pressures of a rise in investment expenditures. A change in v thus results in higher or lower recurring capital gains or losses being generated in the growth process. The level of v as well as changes in the level act as a governor to maintain equilibrium in the markets for securities and current output.

While capital gains or losses created by a rise or fall in security prices may be an important contributory factor to the climate of business expectations in which investment decisions are formed, the main influence of v on investment spending comes through changes in the discount rate implicit in security prices as the present value of an expected future stream of dividend and capital gain income (equations (13) and (13′)). Fisherian firms intent on maximizing share values will take the discount rate implicit in share prices as the required cutoff rate of return on investment. Corporations would expand the rate of investment when $v>1$, and reduce it when $v<1$.[6] With endogenous investment, v would have a long-run equilibrium value lying around unity, where the upper and lower bounds are defined by the degree of managerial independence and imperfections in the market for corporate control.

Table 1 presents empirical estimates of the market value to book or replacement value of corporate equity in the postwar period. As may be seen corporations do not adjust investment spending sufficiently to maintain this ratio equal to one in the short run. But the ratio of market to book or replacement value of corporate equity is a poor measure of the valuation ratio as defined in this paper. Replacement value is estimated by reference to historical costs and depreciation conventions. Market value of equity represents the discounted present value not merely of earnings from real capital as measured in the denominator, but also earnings due to technological change, monopoly rents, research and development, good will, and the skill and experience of the firm's labor force and management. As a result, investment expenditures are not well explained by movements in the ratio of market to replacement value of corporate equity.

III. Differential Wealth-Income Ratios: The Importance of Inequality

Saving represents the amount of current income devoted to asset accumulation. But in balanced growth the latter is simply equal to initial wealth times its rate of growth. If S_H^c represents total household saving out of comprehensive income, the average saving ratio as a proportion of current output may be expressed $s=S_H^c/Y$. If the growth rate of income and wealth are exogenous, the economy's wealth-income ratio depends directly on households' average saving ratio as a proportion of current output, and thus on the functional distribution of income and saving propensities by income form.

---

[6] Since this is a one-good model the problem of finding a price for the capital stock over time does not arise. The absence of a relative price for capital goods prevents instantaneous arbitrage of the valuation ratio equal to unity by corporate investment decisions.

TABLE 1—HISTORICAL ESTIMATION OF THE VALUATION RATIO NONFINANCIAL CORPORATE NET WORTH

A. United States 1900–58 (Billions of Dollars)

Year	Market Value	Book Value	Valuation Ratio[a]
1900	14.7	22.2	66.2
1912	39.3	37.7	104.2
1922	76.1	95.1	80.0
1929	187.5	146.8	127.7
1933	101.7	99.8	101.9
1939	101.5	99.3	102.2
1945A	151.3	164.6	92.0
1945B	148.4	180.5	82.2
1949	148.8	281.7	52.8
1953	220.9	388.2	56.9
1956	386.3	501.3	77.1
1958	470.6	569.0	82.7

B. United States 1947–71

Year	Valuation Ratio	Year	Valuation Ratio	Year	Valuation Ratio
1947	.72	1955	.81	1964	1.20
1948	.60	1956	.77	1965	1.29
1949	.57	1957	.68	1966	1.00
1950	.61	1958	.81	1967	1.07
1951	.59	1959	.92	1968	1.17
1952	.63	1960	.90	1969	1.00
1953	.60	1961	1.07	1970	.87
1954	.75	1962	1.01	1971	.90
		1963	1.17		

C. United Kingdom 1957–61 (Billions of Pounds)

Year	Market Value	Book Value	Valuation Ratio[a]
1957	11.3	26.6	42.5
1958	15.0	28.2	53.2
1959	22.2	30.2	73.5
1960	21.2	32.8	64.6
1961	22.1	35.0	63.1

Sources: **A**: R. Goldsmith and R. Lipsey, Table 25, pp. 126–27; **B**: C. W. Bischoff; **C**: J. Revell, Table 3.7, p. 58.

[a] Shown in percent.

$$s = \frac{S_H^c}{Y} = \frac{gW}{Y} \tag{19}$$

$$\frac{W}{Y} = \frac{s}{g} = \frac{1}{g}\Bigg[s_L + (s_P - s_L)\left(\frac{P}{Y}\right) + (s_G - s_P)\left(\frac{s_B P}{Y}\right) + s_G\left(\frac{gK(v-1)}{Y}\right)\Bigg] \tag{20}$$

As stated the effect of business saving is potentially inflationary, since corporate retained earnings devoted to real capital accumulation become available as capital gains to be spent again by households. Shareholders always have the possibility of "undoing" by realizing as capital gains any forced saving undertaken for them by their corporate managers. The equilibrium wealth-income ratio in balanced growth is simply equal to the average saving ratio divided by the growth rate (equation (20)), and so cannot be higher than the amount households are willing to save and accumulate. Only if households always passively hold all wealth increments received

TABLE 2—WEALTH-INCOME RATIO BY WEALTH CLASS, 1962[a]

	Estimated Number in Millions	Percent of Total Units	Income (Mean in $)	Wealth (Mean in $)	Ratio W/Y	Percent of Wealth	Percent Total of Publicly Traded Stock
All Consumer Units	57.9	100	6,378	20,982	3.290	100	100
Size of Wealth							
Negative	1.0	1.73			—	—	—
Zero	4.7	8.12			—	—	—
$1–999	9.0	15.5	4,121	396	.096	—	—
1,000–4,999	10.8	18.65	5,278	2,721	.516	2	—
5,000–9,999	9.1	15.71	5,970	7,267	1.217	5	1
10,000–24,999	13.3	22.97	6,899	16,047	2.326	18	3
25,000–49,999	6.2	10.70	8,014	35,191	4.391	18	7
50,000–99,999	2.5	4.32	11,302	68,980	6.103	14	14
100,000–199,999	.7	1.21	17,138	132,790	7.748	8	10
200,000–499,999	.5	.86	24,258	300,355	12.382	13	24
500,000 and over	.2	.35	63,508	1,260,667	19.851	22	41

Source: Projector and Weiss, Tables A-8, A-16, A-33, A-36.

[a] Wealth is defined to include the following assets, measured at market value: homes, automobiles, businesses or professions, liquid assets, investment assets, and assets held in trust minus the value of debts secured by these assets, primarily home mortgages. Assets omitted are equity in life insurance, annuities and retirement funds, and personal durables: clothing, furniture, household and sports equipment, jewelry, art collections, coins, stamps, etc.

as capital gains due to corporate retentions ($s_G=1$) will corporate saving behavior never be undone by wealth owners. As shown in equation (18) this would imply that households have no desired finite wealth-income ratio.

As is well known, wealth ownership is highly concentrated. In the United States the most detailed study of upper income and wealth groups has estimated that the wealthiest 10 percent of consumer units accounted for about 65 percent of total household wealth (net of debt secured by these assets and excluding consumer durables other than automobiles) and over 90 percent of all publicly traded stock (see Dorothy Projector and Gertrude Weiss). As shown in Table 2 the wealth-income ratio rises dramatically by income level. The median wealth-income ratio for the lower 85 percent of households is considerably less than one. The composition of wealth portfolios is also sensitive to wealth and income levels. Lower wealth holders own primarily homes, consumer durables, automobiles, and some liquid assets. Top wealth owners hold the bulk of their wealth in the form of business and investment assets, primarily listed and unlisted corporate stock and equity in business and professions.

Since saving represents the demand to accumulate wealth, higher wealth-income ratios necessarily imply higher saving-income ratios if the wealth-income ratio is to be maintained over time. Consider the case of top wealth holders whose income is derived virtually entirely from property, so that their wealth-income ratio is $W/rW = 1/r = s/g$. Since the rate of growth of their wealth is simply the rate of return earned r_c times their saving ratio s_c, the rate of growth of their wealth is $g_c = s_c r_c$. In order to maintain their relative wealth share, their wealth must grow at the rate of growth of the economy. This implies that with rates of return on wealth about double the growth rate of the economy, capitalists would have to have a saving ratio of about 0.5. This in fact approximates the estimated saving ratios for upper income groups (see Simon Kuznets, Irving Friend and Irving Kravis).

This suggests the functional significance of highly concentrated property income and wealth ownership observed in all capitalist societies. If wealth were equally distributed, the economy's wealth-income ratio would be governed by the representative households' saving preferences. With life cycle-consumption behavior and zero terminal wealth bequests, the economy's aggregate wealth-income ratio would be much lower, as limited by the amount households desire to save during their working life to consume during their retirement.[7] Suppose that, after a revolution, all corporate stock were redistributed equally among households. Due to the attempt to realize these claims for consumption purposes, in particular real estate and consumer durables, on the part of many households, equity values would fall drastically. A new equilibrium position would eventually be reached at which the ratio of wealth and capital to income would be much lower, discount and profit rates higher, and investment and growth rates lower.

It is difficult to resist the conclusion that wealth concentration leads to "too much" saving and wealth, relative to what would occur with a more egalitarian distribution of wealth ownership, where household saving preferences would be more equally and therefore more democratically weighted in the market place. What has posterity done for us recently? With a more equal distribution of wealth ownership, a much greater bribe would have to be paid to wealth owners to induce them to refrain from present consumption and maintain any given ratio of wealth to income. When all wealth is concentrated in a few hands, a much smaller bribe is sufficient to induce wealth owners to forego additional present consumption, particularly when they are already consumption leaders.

IV. The Role of Finance in Accumulation: Saving as a Constraint on Investment

In the neoclassical view the rate of profit in the long run was regarded as determined simultaneously by technological relationships, which govern the marginal productivity of capital assets and lie behind the demand for loanable funds, and by time preference and thrift, which govern the saving function and so the supply of loanable funds. The interest rate as the price of hiring purchasing power fell somewhere between the real rental rate for capital goods and the return from holding money, return differentials measuring the risk and liquidity premiums associated with different assets. Interest had to be paid to savers as a reward to waiting, to induce them to defer present consumption and yield up their control of purchasing power to others.

The contribution of Keynesian analysis was to distinguish sharply the return on real assets from the return on fixed-income financial assets, and to weaken the link between them, at least in the short run in which real world economies are continually enmeshed. The interest rate as determined in financial markets reflected the outcome between portfolio preferences for liquidity and existing stocks of financial assets. The functional justification of interest was undermined, since interest was regarded as the reward for foregoing liquidity, noncommitted purchasing power, rather than foregoing present consumption of goods. Saving was regarded as inelastic to the rate of interest, which was strongly influenced by inelastic expectations based on past experience. The interest rate determines the present value of capital goods as the discounted value of their expected future earnings. If this exceeded their minimum supply price, net investment would occur. The rate of investment would depend on the elasticity of supply of the capital goods industries.

---

[7] James Tobin has calculated that with even the relatively low subjective discount rate of 5 percent, the aggregate demand for wealth by life cycle savers becomes negative, as high consumption in the early years is financed by borrowing against future income.

The Keynesian view tends to emphasize the autonomous nature of investment behavior in governing the rate of capital accumulation. The effect of greater investment spending is to generate the higher income, profits, and saving necessary to finance it. "Workers spend what they get, capitalists get what they spend" is Professor Kalecki's well-known aphorism. Investment decisions were made by business in response to anticipations of future profits. Corporations undertook some saving directly in the form of retained earnings, and the rest was generated indirectly by the rise in the level of income. As a result interest became more of a transfer payment, and was no longer viewed as a necessary reward for waiting or abstinence. Keynes was consequently able to visualize the future euthanasia of the rentier as the return on investment fell with increased capital accumulation.

The logic of the present argument in a sense supports both views. The neoclassical conclusions of the control of the rate of accumulation by saver preferences is validated in principle, albeit through a quite different mechanism. It is not necessary to pay interest to induce wealth owners to save and forego current consumption of goods. Business corporations will undertake both the investment and much of the saving necessary to finance it. Nor is interest needed to induce wealth owners to forego purchasing power. All financial claims could easily be invested with liquidity. Rather interest must be paid to induce wealth owners to hold voluntarily and not spend on current consumption the wealth accumulation that results from business investment. Assume that all corporate investments were internally financed. Corporate retentions do represent involuntary saving by corporate owners, undertaken by corporate managers, largely independently of the saving preferences of shareholders as the Keynesians have emphasized. Nevertheless the Keynesian view ignores the fact that holders of equities always have the possibility of "undoing" any such involuntary saving undertaken by their hired managers, simply by realizing for consumption purposes the capital gains that accrue from the investment process.[8]

The mechanism by which this control by wealth owners operates in the present model is through the valuation ratio. Even though wealth owners in the aggregate cannot enjoy more consumption goods than are currently being produced, the attempt to realize wealth and capital gains in order to purchase commodities drives down equity values. Fisherian firms will adjust their investment spending to maximize the equity value of the corporation, and this requires using the rate of discount implicit in share prices as the cutoff rate of return for investment projects. A fall in the valuation ratio raises the cost of capital, reducing the amount of investment undertaken. Such firms adjust investment spending until the valuation ratio is again equal to one, at which point, providing there is an inverse relationship between the return on real capital and the capital-output and capital-labor ratio, the wealth-income and investment-income ratio will have fallen and the output of consumption goods will have increased. The adjustment process takes time; so long as wealth owners cannot consume the existing capital stock directly, reduction comes about by not replacing depreciated capital goods and negative net investment.[9]

---

[8] Capital gains as stated introduce a potentially inflationary element into the economy. Corporate income spent on investment creates income *pari passu* to corporate owners, which can be spent over again for consumption. If wealth owners collectively do not wish to accumulate, they can realize their capital gains and prevent their wealth portfolios from increasing. It is frequently pointed out that the excess of disposable income over earned income due to capital gains rests on an illusion, since if wealth owners collectively decided to realize their gains, the gains would disappear. But it is not always recognized that the same illusion holds for wealth ownership of whatever sort. The stock of wealth represents value relationships in terms of current commodities. If wealth owners were unwilling to hold assets greater than a certain proportion of their annual income, the value of the existing wealth stock, irrespective of how large an inventory of physical assets is represented, must fall to this level. The fact that there is this element of illusion when viewed collectively does not alter the fact that from the viewpoint of the behavior of an individual, one dollar of accumulated wealth or of capital gains income is exactly equivalent to one dollar of earned income in its command over purchasing power. For the individual a high ratio of wealth to income offers potential disposal over goods far in excess of current income, even if this is impossible for the community as a whole.

The same general mechanism also operates for non-Fisherian firms in imperfect capital markets. In this case, rather than having a long-run equilibrium value of one, the valuation ratio would have some upper and lower bound. Within this range corporate decisions to accumulate would be independent of wealth owner preferences and rate of time preference. Only if corporations were completely indifferent to the market price of their stock would they be able to operate completely independently of wealth owner preferences as reflected in financial valuations.

V. Property Income and Economic Rent

While the argument thus leads to the validation of the neoclassical view of control of capital accumulation by saver preferences in principle, in practice as developed above ultimate wealth owners do not choose to undo the accumulation decisions of their corporate managers. If capital gains are always saved, the saving preferences of wealth owners are not a constraint on accumulation, apart from the influence of the valuation ratio in corporate investment. In the short run investment tends to create the saving necessary to finance it, and we are back in a Keynesian world.

Acceptance of such a view would seem to have profound implications for the functional justification of property income. Most of the economy's saving is undertaken by upper income groups, either directly or indirectly by their corporate managers and not undone. Such saving behavior would appear only to be explainable in terms of the satisfaction derived directly from possession and accumulation of wealth.[10] To the extent saving by such groups represents not a conflict between present and future consumption (using up), but the acquisition of long-lived goods yielding current and future satisfaction complimentarily, the traditional functional economic justification of interest as the reward for waiting, for the deferral of present gratification, largely disappears. To the extent that for capitalist households saving does not represent a choice between present versus future satisfaction, because wealth possession and accumulation yield present utility directly, a very large component of economic rent is involved in property income. To the extent asset acquisition for such groups in fact yields substantial services-in-kind, a much smaller or even zero income return would suffice as an inducement to forego absorptive expenditure and hold the economy's capital stock.

---

[9] In a one-good model it would be possible for wealth owners to control the stock of real capital directly in the short run. This would occur if firms bought and sold securities for goods so as to maintain the valuation ratio equal to unity, or any other level above or below one. Household desire to hold less wealth would result in open market purchases of equities by firms in return for commodities, so that households could consume some of the existing capital stock directly. Conversely household desire to hold more wealth would result in new issues of equities by firms in exchange for commodities, so that real capital accumulation would occur *pari passu*. With such a perfectly accommodating financial system aggregate demand would always equal full employment aggregate supply, since excess demand for commodities would never appear.

[10] For a criticism of life cycle-saving theories which emphasizes the distinction between labor and property income concerning the bequest motive for saving, see the author (1973b).

This does not of course imply that even with a zero reservation price of saving, a zero rate of return is possible or even appropriate, under capitalism or socialism. Financial returns like all prices have both an allocative and distributive effect (see James Meade). Some positive cost of finance is required to allocate scarce resources efficiently between alternatives. But from a distributional viewpoint property income is very largely a transfer payment from workers to owners, a quasi rent like the return to ownership of land, which apart from saving incentives has little economic justification as a reward to ownership per se. The fact that capital goods are productive cannot in itself be used to justify the assignment of their value contribution to their private owners, particularly if no economic sacrifice or burden or function is involved in the ownership and accumulation of such wealth. One of the unattractive apologist implications of conventional economic analysis is that it serves to elevate rents, profits, interest, and all property income to the level of respectability of wages (see Joan Robinson). Moreover property income, the quasi rents paid to the owners of claims against the product of past labor inputs, can only under very restrictive assumptions be regarded as measuring the marginal productivity of any "mythical" current capital stock.[11]

Consider the limiting case in which corporations rely entirely on internal finance for investment expenditures ($I=s_BP$) and households save all capital gains received from stock ownership ($s_G=1$, $s_L=s_P=0$). In such a world there would be no financial constraint on investment: saving by corporations and households would always be exactly sufficient to finance corporate investment spending. The supply of saving function would be identical to the demand for investment function. In such a world what forces would determine the profit rate and limit the level of net investment? Irrespective of the volume of investment chosen, there would never be a problem of financing or of insufficiency of aggregate demand. An increase in the share of investment would require an increase in retained earnings and, assuming the payout ratio is constant, a rise in profits and so in the ratio of prices to costs. Increased investment spending thus necessarily involves lower consumption spending and a lower level of real wages w/P.

If wages are regarded as determined by a struggle over distributive shares between workers and capitalists, Robinson has argued that there is some minimum level of real wages at which money wages tend to rise rapidly as workers are determined and able to resist further deterioration in their real wages. This "inflation barrier" sets a limit on the ability of firms to increase profit margins to finance a higher level of investment and so limits the investment income ratio and the rate of accumulation. The Knightian argument is that in a world of uncertainty investment would never be carried beyond the point where the expected rate of profit on new investment was sufficiently high to compensate for the irreducible degree of uncertainty involved in business enterprise. The degree of risk aversion on the part of firms thus sets a floor to the required rate of return to investment. Competition will also set a limit to the rate of investment in such a world. Firms can increase investment only by increasing retained earnings and, after payouts have been reduced to their minimum value, by raising the ratio of prices to costs. As a result they will lose business to firms that are content to grow more slowly and therefore can charge lower price markups over cost, since such firms do not require as large a volume of retained earnings.

---

[11] The expected rate of profit on new investment must be clearly distinguished from the observed rate of profit on capital P/K. Profits received on existing assets consist of a series of rents and quasi rents. With technological change the quasi rents earned on older capital goods decline with age until they are scrapped. The value of the existing capital stock cannot in general be determined independently of the profit rate. As a result the observed ratio P/K is of little analytical meaning outside of a golden age. As Mario Nuti states, "The ideological role of the value of capital is that of breaking the actual direct link between the time pattern of labor inputs and the time pattern of output in which any technology can be resolved, and establishing instead a relationship between current output and current labor" (p. 54).

The present argument suggests an additional mechanism which operates to limit the rate of investment and set a floor to the rate of profit. As shown in equation (13) the long-run rate of return on corporate stock (when $i=0$ and $dv=0$) cannot fall below the dividend yield plus the growth rate of corporate earnings (assets). At the limit as share prices become indefinitely large the dividend yield approaches zero, and the return on equities approaches the growth rate. But this return on equities operates to govern the rate of profit so long as firms are concerned about the price of their shares. As shown above (equation (13′)), when all firms operate to maximize share prices so that $v=1$, the rate of profit (P/K) will be equal to the discount rate r.

In capitalist economies the return on property claims cannot be reduced below the growth rate of the future stream of income to which the owner of the claim is entitled. As returns fall towards this level the market values of property claims, as the discounted value of future income streams, become indefinitely large, as does the ratio of comprehensive income to output.[12] Thus even in a world in which by assumption all property income represents economic rent, and rates of return are totally functionless from the viewpoint of inducing saving, it would not be possible to eliminate property income, nor to reduce the return on wealth ownership below the growth rate, even though no income return is required to induce wealth owners to forego present consumption and bribe them to accumulate.

A final implication of the analysis is that while the profit rate and the rate of return on equities indeed rests on three legs, only one of these, the productivity of new investment, was recognized by Böhm-Bawerk. The other two are the growth rate of the economy and the proportion of property income which is saved and invested. The profit rate must be sufficiently high to finance the level of investment consistent with the capital-labor ratio and rate of growth of effective labor. The higher the proportion of profits saved (s_B, s_G, s_P, s_L), the lower the profit rate. But the aggregate saving ratio depends more on institutional factors, in particular wealth distribution, than on time preference.

In a modern market economy the volume of household saving is closely dependent on the volume of business investment, depending on the degree of wealth concentration. While lower and middle income individuals may adjust their present and future consumption plans in response to a change in the rate of return on wealth, they are not "representative" savers, since their margin is not significant in governing aggregate saving behavior.[13] Equity returns and profit rates would be positive in growing economies even if the reservation price of savers were zero and all property income represented pure economic rent.[14] To conclude that the rate of profit on capital and the rate of return on wealth is determined by individual time preference represents a colossal inversion of causality.

---

[12] As noted above in a growth context the reduction of property returns towards the growth rate of the economy, and not to zero, is sufficient for the long-run euthanasia of the rentier. Rentier saving ratios would then have to approach 100 percent in order to prevent their relative share of total wealth from declining to zero in the long run.

It is concentration of property ownership in addition to the capitalist ethic that make capitalism as a system such a powerful engine of accumulation. Paradoxically, by ensuring that a large proportion of property income is saved, concentration of wealth serves in part to condone the institution of private property itself. In Keynes' famous double bluff passage he observed that capitalists, though they do no work, "were allowed to call the best part of the cake theirs, and were theoretically free to consume it, on the tacit underlying condition that they consumed very little of it in practice." Or as another famous critic of capitalism has put it, "Accumulate! Accumulate! That is the law of Moses and the Prophets." But one price is the very considerable inequality of income and opportunity, which could be reduced by the differential taxation of property income, or by the collective ownership of property claims to society's real wealth. This policy conclusion is reinforced by the first bluff, that it is the abstention of workers and not of capitalists that provides the surplus for accumulation.

REFERENCES

J. Arena, "Postwar Stock Market Changes and Consumer Spending," *Rev. Econ. Statist.*, Nov. 1965, *47*, 379–91.

M. Bailey, "Capital Gains and Income Taxation," in A. Harberger and M. Bailey, eds., *The Taxation of Income from Capital*, Washington 1966.

K. B. Bhatia, "Capital Gains and the Aggregate Consumption Function," *Amer. Econ. Rev.*, Dec. 1972, *62*, 866–79.

C. W. Bischoff, Yale University, unpublished estimates.

P. David and J. Scadding, "Private Savings: Ultra Rationality, Aggregation and 'Denison's Law'," *J. Polit. Econ.*, Mar. 1974, *82*, 225–50.

M. Feldstein, "Taxes, Corporate Dividend Policy and Personal Savings: The British Postwar Experience," disc. pap. no. 20, Harvard Inst. Econ. Res., Aug. 1972.

I. Friend and I. Kravis, "Consumption Patterns and Permanent Income," *Amer. Econ. Rev. Proc.*, May 1957, *47*, 536–55.

——— and ———, "Entrepreneurial Income, Saving and Investment," *Amer. Econ. Rev.*, June 1957, *47*, 269–301.

R. Goldsmith and R. Lipsey, *Studies in the National Balance Sheet of the United States*, Vol. 1, Princeton 1963.

R. Holbrook and F. Stafford, "The Propensity to Consume Separate Types of Income," *Econometrica*, Jan. 1971, *39*, 1–22.

N. Kaldor, "Alternative Theories of Distribution," *Rev. Econ. Stud.*, June 1956, *23*, 83–180.

———, "Marginal Productivity and the Macro-Economic Theories of Distribution," *Rev. Econ. Stud.*, Oct. 1966, *33*, 309–19.

L. Klein and A. Goldberger, *An Econometric Model of the United States, 1929–52*, Amsterdam 1955.

J. M. Keynes, *The Economic Consequences of the Peace*, London 1919.

F. Knight, "Neglected Factors in the Problem of Normal Interest," *Quart. J. Econ.*, Feb. 1916, *30*, 279–310.

---

[13] For an early expression of the position that the rate of interest is not governed by time preference, see Frank Knight. He argued that "the thing desired is possession," and not future consumption, that "the motivation is not economic in the ordinary sense but sociological" (p. 286). He detailed a number of reasons why net saving was carried beyond the individual lifetime of members of the social class from which capital accumulations actually come.

[14] The fact that asset accumulation continues in the present inflation, when real returns on fixed income financial assets are negative, is an indication of the interest inelasticity of the supply of saving, at least in the short run, and the extent of economic rent in property income.

J. Kregel, *Rate of Profit, Distribution and Growth: Two Views,* Chicago 1971.

S. Kuznets, *Shares of Upper Income Groups in Income and Saving,* New York 1953.

J. Meade, *Efficiency, Equality and the Ownership of Property,* London 1964.

F. Modigliani, "Monetary Policy and Consumption: Linkages via Interest Rate and Wealth Effects in the FMP Model," in *Consumer Spending and Monetary Policy,* Fed. Res. Bank Boston, *Monetary Conference Series,* No. 5, June 1971.

B. J. Moore, (1973a) "Some Macroeconomic Consequences of Corporate Equities," *Can. J. Econ.,* Nov. 1973, *6,* 529–44.

———, (1973b) "A Reexamination of Life Cycle Theories of Saving," work. pap., Wesleyan Univ. 1973.

M. Nuti, "Capitalism, Socialism, and Steady Growth," *Econ. J.,* Mar. 1970, *80,* 32–57.

D. Projector and G. Weiss, *Survey of Financial Characteristics of Consumers,* Washington 1966.

J. Revell, *The Wealth of the Nation, The National Balance Sheet of the United Kingdom, 1957–61,* Cambridge 1967.

J. Robinson, *Economic Heresies,* New York 1971.

J. Stein, "Monetary Growth Theory in Perspective," *Amer. Econ. Rev.,* Mar. 1970, *60,* 85–106.

J. Tobin, "Life Cycle Saving and Balanced Growth," in *Ten Economic Studies in the Tradition of Irving Fisher,* New York 1967, 231–56.

Constraints on Government Investment Opportunities and the Choice of Discount Rate

By David F. Bradford*

In his well-known paper on the choice of discount rate for public investment decisions, William Baumol noted the apparent impossibility of reconciling the arguments in favor of using a pure time preference rate with those in favor of using the marginal rate of return to private investment, the "opportunity cost" of funds. On the one hand it would seem there is a clear opportunity for welfare gain in undertaking an investment with rate of return in excess of a social time preference rate, however determined. On the other hand, it would clearly be possible to do even better by leaving the resources in the private sector if the rate of return on private investment exceeded the return on the government project.

The way out of this dilemma which had already been expressed by Kenneth Arrow (and was subsequently elaborated upon by Arrow and Mordecai Kurz), is to make explicit the optimization problem, including its constraints, confronting the government. Assuming there is a solution to this optimization problem, there is, correspondingly, a "correct" discount rate (or, more generally, term structure of discount rates) applicable to the project.

In this paper I explore the practical relevance of one of the more remarkable conclusions of Arrow's paper, which can be roughly stated as follows: if capital market imperfection takes the form of a fixed marginal propensity to save out of private income (independent of the rate of return), the optimal government investment policy in the long run is to invest to the point where the marginal rate of return on government capital equals the marginal rate of social time preference (whether or not derived from individuals' rates of time preference), *regardless of the rate of return on private capital.*

Whereas Arrow considered the problem of optimally setting the level of a homogeneous public capital stock, I address the question of whether a project producing certain output flows should be accepted or rejected, where the decision is assumed not to affect any of the relevant marginal rates of transformation or substitution, which are regarded as constant. It proves possible to use a very simple line of argument to show why and under what conditions the analogue of the Arrow proposition holds and to indicate how modification of the conditions of the problem change the public investment criterion. The policy of discounting according to a pure time preference rate turns out to be surprisingly robust to plausible variations in the parameters describing the underlying (second best) optimization problem.

All of the results, it should be stressed, describe second best policies. That is, they depend upon the existence of constraints on government investment opportunities which are "artificial" in the context of the economic model, although possibly neither unrealistic nor unreasonable as political devices. These are the constraints which restrict government investment to certain classes of activities. Because of such restrictions, apparently attractive returns in the private sector may not represent opportunity costs for the government at all, or at least not directly. When the private investment opportunities are not also government investment opportunities, private rates of return become irrelevant to government choices. This is the key to the argument.[1]

---

* Princeton University. An earlier version of this paper appeared as Research Memorandum No. 67, Institute for Advanced Studies, Vienna, May 1972. My debt to the work of K. J. Arrow and M. Kurz is great, and through them to the work of S. A. Marglin and O. Eckstein. I would like to acknowledge and recommend as well a paper by Martin Feldstein which uses a similar approach, although I choose to stress different aspects of the problem.

Section I contains the basic model and the most important conclusions. Briefly, the argument focuses on the comparison between the amount of reduction in private capital formation caused by the shift of one dollar into the public sector and the increase in private capital formation induced by an increase of one dollar in the value of government services. If capital market imperfection causes the value of the extra capital obtainable by a reduction of one dollar in consumption to exceed one dollar (shadow price of capital in excess of unity), a given government investment becomes the more attractive the smaller the private investment-discouraging effect of its financing and the greater the private investment-encouraging effect of its output flow. For the special case in which these private investment effects are at the same rate per dollar, and assuming the shadow price of capital to be constant, a public investment with rate of return in excess of the rate of time preference should be undertaken. This would, under the assumption of capital market imperfection, imply government takeover of private investment, a rather radical result ruled out by the assumed limitation of the government's sphere of operation. More modestly, the result implies (for the noted special case of the parameters) that the government should use the rate of time preference in discounting its available investment opportunities.

The remaining sections are principally concerned with amplifications and extensions required to move the analysis toward practical applicability. In Section II, I discuss the procedure for calculating the shadow price of capital and present plausible illustrative magnitudes. In Section III the model is extended to recognize many different private investment sectors, a refinement principally motivated by the view that a corporation income tax leads to different (social) rates of return on different forms of private capital. In Section IV the case of multiperiod projects is considered. The most important contribution of this exercise is to emphasize that the basic discount rate in the government's calculations is the time preference rate. The private-investment parameters here determine factors by which costs or benefits are to be inflated or deflated. While for the case of one-period investments these effects can be equivalently expressed as variations in the discount factor, for multiperiod investments this is not so. For investments of any significant duration the distinction is crucial. This point is used in Section V in which, after commenting on the prospects for detailed application of the model, I return to the thesis that discounting at a pure time preference rate is a defensible rule of thumb for government choices.

Some final introductory remarks: in this paper the features of the investment problem arising from uncertainty are given no explicit attention, even though these are presumably relevant to explaining the simultaneous existence of different rates of return. For our purposes here it is sufficient to view all rates of return and time preference as certain. It is implicitly assumed, furthermore, that it makes sense for the government to base its investment decisions on preferences about *aggregate* consumption flows. That is, it is assumed that either the intratime period income distributional effects of investment choices are unimportant, or mechanisms are at hand to adjust the income distribution to any desired extent within each time period.

---

[1] In a previous paper, I showed how constraints on government actions affected choice criteria in the presence of multiple rates of time preference contemporaneously held by the members of society. In this paper I am concerned with a different problem. The source of the trouble here is multiple rates of return on investment. It arises even if all members of society (or some designated welfare evaluator) are in agreement about the appropriate rate of substitution among consumption in different time periods.

I. The Basic Argument in a Simple Model

The essence of the analysis is adequately expressed in context of a simple model, in which it is assumed there is only one kind of private capital and all the yield from an investment takes place in the immediately following period. In such a model, in fact, one does not need to make a distinction between capital stock and investment flow. In each period the whole capital stock is up for grabs, and may be consumed or invested.

Let c_t stand for the level of aggregate consumption in period t. It is assumed that the government acts according to a preference structure over time streams of consumption. For our purposes it is not necessary to establish the basis—for example, the time preferences of the households at each moment—on which the government determines the relative desirability or "social value" of alternative consumption paths. From such a structure emerges naturally a sequence of "social rates of time preference" i_t, corresponding to each sequence of consumption levels c_t: $1+i_t$ is simply the social marginal rate of substitution of c_{t+1} for c_t.

Assume that investment in private sector capital has a marginal one-period rate of return r_t in period t. That is, $1 invested in period t increases the total number of dollars available in period t+1 by $(1+r_t)$ dollars. An increase in the private capital stock at any time may be expected to have an impact on the entire future consumption stream, since some of the proceeds are likely to be consumed in the next period, while a further part will be reinvested to affect consumption still farther in the future. Obviously, it may be a difficult task to identify the "perturbation" of the consumption stream resulting from a $1 change in private capital at time t, but let us imagine that this has been done, and let v_t stand for the discounted value of the sequence of consumption changes, where the discount factors are derived in the usual manner from the sequence i_t of social rates of time preference.

To clarify what v_t is, imagine that a unit of time t capital drops like manna from heaven, affecting the whole stream of consumption starting at time t+1. In this new consumption stream, by how much at most could one reduce the amount of consumption in period t, and still have a consumption stream as valuable (according to the government preferences) as the original, pre-manna, consumption stream? The answer is v_t, and in this sense, v_t is the social value of a unit of private capital at time t. Note carefully that v_t is measured in units of time t consumption. Because of their analytical function, I shall adopt Feldstein's practice of calling the v_t simply the "shadow prices" of private capital.

Now let us suppose that the government is contemplating undertaking an investment in period t which has a rate of return ρ_t. That is, $1 invested in the government opportunity in question will make available an increase of $(1+\rho_t)$ dollars in period t+1. The question now is, under what conditions would an incremental unit of the government investment lead to a better time stream of consumption (from the social point of view). The answer is made difficult by the fact that, through its influence on private capital formation in periods t and t+1, the government investment decision affects the entire future consumption sequence. However, having equipped ourselves with the sequence of social values of capital v_t, we are in a position to evaluate these indirect effects more easily.

Let a_t be the amount by which private capital formation is decreased as a result of the financing of an additional dollar of government investment. Under the usual full employment assumption, $(1-a_t)$ is the corresponding loss in period t consumption.[2] For some forms of finance we would expect to find $a_t=s_t$, the marginal propensity to save out of disposable income, although clearly we shall wish to associate different a_t with different techniques of financing government investment.[3]

Let α_t be the amount by which private capital in period t is increased as a consequence of an increase of $1 in the output of the government sector in period t. Whereas a_t shows the *reduction* in private capital formation caused by an increase of one unit in financing, α_t shows the *increase* in private capital formation caused by a unit flow in the other direction in the form of government services, measured in dollars. For the case in which revenue is raised by direct taxation of consumer citizens and in which the implicit income from the government project is treated exactly like ordinary, after-tax income, $a_t=\alpha_t=s_t$. (This is the Arrow-Kurz assumption.) However, we must admit the possibility that these three parameters differ.

TABLE 1—THE EFFECT OF UNDERTAKING AN ADDITIONAL ONE DOLLAR OF GOVERNMENT INVESTMENT AT TIME t

	Period t	Period t+1
Change in Consumption	$-(1-a_t)$	$(1+\rho_t)(1-\alpha_{t+1})$
Change in Private Capital Formation	$-a_t$	$(1+\rho_t)\alpha_{t+1}$

It is now a simple matter to tote up the various effects of raising an additional dollar in period t to finance government investment. Table 1 shows the impact on consumption and capital formation in periods t and t+1. We know that the value in period t consumption units of $1 of private investment is v_t. Hence the effective loss in period t consumption as a result of the government investment is $1-a_t+a_tv_t$ dollars. Similarly, the value of the increase in consumption and private investment in period t+1, expressed in period t+1 consumption units, is $(1+\rho_t)\cdot(1-\alpha_{t+1}+\alpha_{t+1}v_{t+1})$. Discounting the effective t+1 consumption increase to period t at the social rate of time preference i_t, we obtain an expression for the net gain from the whole transaction:

$$(1)\quad -(1-a_t+a_tv_t)+\left(\frac{1+\rho_t}{1+i_t}\right)(1-\alpha_{t+1}+\alpha_{t+1}v_{t+1})$$

By definition of i_t, we shall attain a more valued consumption stream if we undertake government investment so long as expression (1) is positive, a condition expressed by inequality (2),

$$(2)\quad \frac{1+\rho_t}{1+i_t}\geq\frac{1+a_t(v_t-1)}{1+\alpha_{t+1}(v_{t+1}-1)}$$

which determines by how much, if at all, the rate of return ρ on the government investment must exceed the social time preference rate i in order to make the investment worthwhile.

---

[2] It would not be difficult to recognize the possibility of underemployed resources, and hence to include explicitly separate parameters for the investment and consumption effects here, but it would clutter up the argument.

[3] The use of the general approach taken here to analyze the choice of financing method is the principal subject in Feldstein's paper.

A. *Some Important Special Cases*

It is useful to consider the implications of condition (2) as a government investment criterion in special cases corresponding to various assumptions about the coefficients a_t, α_t, and v_t. The effect of a rising or falling shadow price of private capital works in the same direction under all meaningful assumptions about the other coefficients. A glance at conditions (2) will show that increasing v_{t+1}, holding v_t constant, tends to raise the denominator on the right-hand side and hence to reduce the critical rate at which ρ_t becomes an acceptable rate of return on the government project. The effect of decreasing v_{t+1}, holding v_t constant, is in the opposite direction. We shall interpret this result below. For the moment we simply note that to consider the consequences of varying the other parameters it is sufficient to look at the cases for which $v_t = v$, a constant. There are four extreme cases of special interest.

Case 1: $a_t = \alpha_{t+1}$. An additional dollar's worth of output from a government project causes the same increment in private capital formation as does a reduction of $1 in the amount raised through the financing instrument corresponding to which a_t is defined. This case is the rather remarkable one distinguished in the introduction. Substituting into condition (2) we see that the acceptance condition for the public investment becomes

$$\frac{1 + \rho_t}{1 + i_t} \geq \frac{1 + a_t(v - 1)}{1 + a_t(v - 1)} = 1 \tag{3}$$

This says that *government investments should be accepted as long as the marginal rate of return exceeds the time preference rate.* The surprising aspect of this case is that the rate of return in the private sector, sometimes called *the* opportunity cost of funds, does not enter the calculation at all. The reason is that by our assumption that $\alpha_{t+1} = a_t$, we have assured that for every dollar of reduced investment in period t, the government project puts back $(1+\rho_t)$ dollars of increased investment in period t+1. The issue then is, how much must we increase investment in period t+1 to compensate for a loss of one unit in period t. Since $v_t = v_{t+1}$, the answer is $(1+i_t)$ dollars.

Case 2: $a_t = 1$, $\alpha_{t+1} = 0$. Under these assumptions, which are often made, at least implicitly, *all* of the resources for the project come from investment and the output induces no increase in private investment. Here the usual argument is that a project is acceptable if and only if it has a rate of return greater than or equal to that in the private sector. According to the present analysis this is generally not strong enough. Condition (2) becomes

$$\frac{1 + \rho_t}{1 + i_t} \geq v_t \tag{3'}$$

Under the conditions in which v_t, i_t, r_t, and s_t (which we have called the marginal propensity to save) are all constant, we shall see in Section II that $v_t > \gamma \equiv (1+r)/(1+i)$, indicating that the government project must have a return rate in excess of r to pass muster under the assumptions of Case 2. The reason for this is simply that the government project does not generate the favorable repercussions on future capital formation which the private investment does.

Case 3: $a_t = 0$, $\alpha_{t+1} = 1$. In this case all resources for the project come from consumption and all yield is converted into private capital. Condition (2) becomes

$$\frac{1+\rho_t}{1+i_t} \geq \frac{1}{v_t} \tag{3''}$$

For obvious reasons this case would allow as desirable projects for which rate of return is actually below the time preference rate, and a fortiori below the rate of return in the private sector.

Case 4: A Two-Period World. Many analyses of the public investment problem have employed a two-period model, with the second period taken to represent "the future." For some purposes this is a satisfactory procedure, but in the present context it amounts to a very special case of our general analysis, and one with no particular claim to policy relevance. Since in a two-period world it makes no sense to speak of investment in the second period, we have $\alpha_{t+1}=0$. The other special aspect of the two-period world is the ease with which one can calculate the value of v_t. For with no capital formation in period t+1, a unit increment to period t capital must generate $(1+r_t)$ extra units of consumption. The period t consumption value of this, v_t, is $(1+r_t)/(1+i_t)$. With these special assumptions, condition (2) becomes

$$\rho_t \geq (1-a_t)i_t + a_t r_t \tag{3'''}$$

i.e., the condition that the yield on the government investment must exceed a weighted average of the rate of time preference and the private investment yield, the weights being the proportions in which the resources are taken out of consumption and private investment, respectively. Note the very special assumptions underlying this conclusion: no future capital formation consequences of increments in *either* government or private investment.[4]

B. *Summing Up the Analysis of the Simple Model*

The essential conclusions to be drawn from the simple model are obtained from an examination of condition (2). It says that the required yield on a government investment project exceeds the rate of time preference if the dollar reduction in current private investment per dollar withdrawn to finance the government project exceeds the dollar increase in private investment per dollar of value of the project's output $(a>\alpha)$ (assuming $v>1$). The required yield on the government project falls below the time preference rate if the inequality is reversed. The extent of the divergence in each case depends upon v; the larger is v, the larger is the divergence.

As has been pointed out, a tendency for v to grow with time favors government investment, *ceteris paribus*. The reason for this is that the government investment serves in part to *shift private investment toward the future*. The more rapidly v is rising, the less future private investment do we need to offset the loss of any given amount of current investment. By the same line of reasoning, a tendency for v to fall with time raises the minimum acceptable rate of return on government projects.

Although there remains much to be said in the following sections about how one might estimate the values of the various parameters, especially v_t, and about how the simple model can be generalized, the most important theoretical points have already been established. For the analysis has shown how it is possible to reach a definite resolution of the difficulty described by Baumol and referred to in the introductory section. The solution generally lies, interestingly enough, not on either horn of the dilemma, time preference or private productivity discounting, and may not even lie between the two apparent extremes.

---

[4] Peter Diamond derived this same condition, at the same time stressing the limitations of a two-period model. Agnar Sandmo and Jacques Drèze also obtain this formula in a two-period model.

The second best character of these conclusions, resulting from constraints on the government's investment opportunities, must be strongly stressed. How, it will be asked, can it possibly make sense for the government to invest in a project with a yield of 5 percent when there are projects available in the private sector yielding 10 percent? The answer is that this can make sense ***only if the private sector investment is not also an investment opportunity for the government.*** In our model the government participates in the private sector investment opportunity only indirectly, via its choice of public sector investments (and via its choice of financing instruments). By taking into account in advance the response of private sector investment to government actions it is possible to evaluate the entire stream of consequences of a particular choice. Because of the limited set of investment opportunities considered appropriate for the government, and because of the divergence between private and social value of private capital, it can follow that it is better to undertake the 5 percent project than not to undertake it, which is the real choice. The apparent 10 percent opportunity cost is no such thing, since the investment in question does not represent an "opportunity" for the government at all.

II. Calculating the Shadow Prices

By keeping track of the string of consumption consequences of adding a unit of private capital to the system at time t, and discounting them to time t, we arrive at an expression for v_t:

$$(4)\quad v_t = \gamma_t\left[(1 - s_{t+1}) + \sum_{k=2}^{\infty}(1 - s_{t+k})\prod_{j=1}^{k-1}\gamma_{t+j}s_{t+j}\right]$$

where

$$(5)\qquad \gamma_{t+j} = \frac{1 + r_{t+j}}{1 + i_{t+j}}$$

For the special case of $\gamma_{t+j}=\gamma$, and $s_{t+j}=s$, constants, (4) becomes

$$(6)\qquad v_t = (1 - s)\gamma\sum_{k=0}^{\infty}(s\gamma)^k$$

or

$$(7)\qquad v_t = \frac{(1 - s)\gamma}{1 - s\gamma}$$

assuming $s\gamma<1$, as necessary for convergence.

Since in most cases of interest $r>i$, and hence $\gamma>1$, we find $v_t>1$; an extra dollar invested at time zero generates a stream of consumption changes worth more than \$1. Note further that generally $v_t>\gamma=(1+r)/(1+i)$. We may check that for the special case $i=r$, $v_t=1$.

A. *A Difference Equation Formulation*

Expression (4) can be regarded as a difference equation in the shadow price of private capital, which can be derived directly. An increment of investment at time t generates extra consumption at t+1 equal to $(1+r_t)(1-s_{t+1})$, and extra investment equal to $(1+r_t)s_{t+1}$. The latter is equivalent to $(1+r_t)s_{t+1}v_{t+1}$ units of period t+1 consumption, by definition of v_t. Finally, since c_{t+1} units of t+1 consumption are equivalent in social value to $c_{t+1}/(1+i_t)$ units of time t consumption, by definition of i_t, we conclude that \$1 extra invested at time t has a value

$$(8)\quad v_t = \frac{(1 + r_t)(1 - s_{t+1}) + (1 + r_t)s_{t+1}v_{t+1}}{1 + i_t} = (1 - s_{t+1} + s_{t+1}v_{t+1})\gamma_t$$

For the constant coefficient case, $\gamma_t=\gamma$ and $s_t=s$, (8) has steady-state solution $v_t=(1-s)\gamma/1-s\gamma$ to which convergence is

Table 2—The Shadow Price of Private Capital Under a Range of Assumptions About Parameter Values

	$s=.10$			$s=.20$			$s=.30$		
r	.05	.10	.15	.05	.10	.15	.05	.10	.15
i									
.02	1.03	1.09	1.14	1.04	1.10	1.16	1.04	1.12	1.19
.05	1.00	1.05	1.11	1.00	1.06	1.12	1.00	1.07	1.14
.08	.97	1.02	1.07	.97	1.02	1.08	.96	1.03	1.10

assured if $s\gamma<1$. If the system has always had these constant coefficients, the shadow price must be constant at the steady-state value. Note that one can use (8) to compute v_t under more general conditions by making assumptions, for example, about its rate of growth.

B. *Illustrative Numerical Values of the Shadow Price of Capital*

The simple formulation which results from assuming the various coefficients to be constant makes it easy to get some feel for the magnitudes which might be involved in application. The cells of Table 2 contain the shadow prices of private capital implied by various combinations of values of the three parameters, s, i, and r, chosen to bracket the empirically plausible range. The range of v is from 0.96 ($s=.20$, $r=.05$, $i=.08$) to 1.19 ($s=.30$, $r=.15$, $i=.02$).

III. Generalizing to Many Private Investment Sectors

There are a number of ways in which the simple model can be generalized, and it would probably be of little use at this point to produce a catalogue of possibilities. However, in view of the previous treatment of this topic by Baumol and others, we should consider explicitly the possibility of more than one private sector. It is customary by now to think of private investment as being of two types, corporate and noncorporate, where decisions in the former sector are influenced by the corporation income tax. Hence, we start by considering a two-private-sector model.

Let a_t^1, then, represent the fraction of an incremental dollar raised to finance government activities which comes out of noncorporate capital formation. Let a_t^2 be the corresponding value for corporate capital. Define α_t^1 and α_t^2 in the analogous manner, indicating the amounts of capital formation in sectors 1 and 2 induced by a $1 increase in government output in period t. Let v_t^1 and v_t^2 be the appropriate shadow prices of capital in the two sectors, given their respective rates of return, r_t^1 and r_t^2. Table 3 shows the effect of undertaking an additional $1 of government investment at time t in this model world. By a line of reasoning parallel to that leading to condition (2), the criterion for acceptance of the government investment can be expressed as

$$\text{(9)}\quad \frac{1+\rho_t}{1+i_t} \geq \frac{1+a_t^1(v_t^1-1)+a_t^2(v_t^2-1)}{1+\alpha_{t+1}^1(v_{t+1}^1-1)+\alpha_{t+1}^2(v_{t+1}^2-1)}$$

The interpretation of condition (9) runs along the same lines as our previous interpretation of condition (2), and presents no special difficulties. Of particular interest is the case $a_t^1=\alpha_{t+1}^1$, $a_t^2=\alpha_{t+1}^2$ in which dollars taken out of the private sector by the financing methods and implicit dollars put

TABLE 3—THE EFFECT OF AN ADDITIONAL ONE DOLLAR OF GOVERNMENT INVESTMENT WHEN THERE ARE TWO PRIVATE INVESTMENT SECTORS

	Period	
	t	t+1
Change in Consumption	$-(1-a_t^1-a_t^2)$	$(1+\rho_t)(1-\alpha_{t+1}^1-\alpha_{t+1}^2)$
Change in Noncorporate Capital Formation	$-a_t^1$	$(1+\rho_t)\alpha_{t+1}^1$
Change in Corporate Capital Formation	$-a_t^2$	$(1+\rho_t)\alpha_{t+1}^2$

into the private sector in the form of yield on the government investment are divided among consumption and investment in the two sectors in the same proportions. In this case, if the shadow prices of capital are constant, condition (9) becomes simply

$$\frac{1+\rho_t}{1+i_t} \geq 1$$

The government project's yield rate need be only as high as the social rate of time preference.

A. *Calculating the Two Shadow Prices*

If we assume constant propensities to save/invest, constant rates of return, constant social rate of time preference, and constant values of the two shadow prices, v^1 and v^2, we can calculate the latter without particular difficulty. Again, by a derivation analogous to that leading to equation (8), we conclude that v^1 and v^2 must satisfy

$$(10)\quad v^1 = \frac{1+r^1}{1+i}\left[1 + s^1(v^1-1) + s^2(v^2-1)\right]$$

$$v^2 = \frac{1+r^2}{1+i}\left[1 + s^1(v^1-1) + s^2(v^2-1)\right]$$

Define γ^1 and γ^2, analogous to γ in the one-sector analysis. Then, assuming $1-\gamma^1 s^1 - \gamma^2 s^2 \neq 0$, (10) implies that the shadow prices are proportional to γ^1 and γ^2:

$$(11)\quad \begin{bmatrix} v^1 \\ v^2 \end{bmatrix} = \left(\frac{1 - s^1 - s^2}{1 - \gamma^1 s^1 - \gamma^2 s^2}\right)\begin{bmatrix} \gamma^1 \\ \gamma^2 \end{bmatrix}.$$

Under the assumption sometimes made that the social rate of time preference is equal to the private rate of time preference and that this is in turn equal to the rate of return in the noncorporate sector, $r_t^1 = i_t$, i.e., $\gamma^1 = 1$. Notice that because some of the throwoff from each sector is invested in the other, *even in this case* generally $v^1 > 1$.

If we increase the number of private investment sectors beyond two, the government investment test changes from (9) to

$$(12)\quad \frac{1+\rho_t}{1+i_t} \geq \frac{1 + \sum_{j=1}^{m} a_t^j (v_t^j - 1)}{1 + \sum_{j=1}^{m} \alpha_{t+1}^j (v_{t+1}^j - 1)}$$

where m is the number of sectors. Again, note the simple and symmetric way in which the various private sectors enter criterion (12). For the case in which the various coefficients and the shadow prices are constant over time we can calculate them in the manner of expression (11) from

$$(13)\quad \begin{bmatrix} v^1 \\ v^2 \\ \cdot \\ \cdot \\ v^m \end{bmatrix} = \left(\frac{1 - \sum_{j=1}^{m} s^j}{1 - \sum_{j=1}^{m} \gamma^j s^j}\right)\begin{bmatrix} \gamma^1 \\ \gamma^2 \\ \cdot \\ \cdot \\ \gamma^m \end{bmatrix}$$

where $\gamma^j \equiv (1+r^j)/(1+i)$. Here, as in the two-sector example, shadow prices v^i are all in the *same proportion* to (and in the normal case larger than) the coefficients γ^i, which represent the social-time-preference discounted value of marginal returns in sector i.

IV. Treatment of Multiperiod Returns

Thus far the analysis has been carried on as though the government were always in a position of choosing an investment requiring inputs of \$1 in period t and producing its entire output $\$(1+\rho_t)$ in period t+1. This permitted us to express the investment criterion in terms of the relationship between a well-defined rate of return (essentially, the one-period internal rate of return) and corresponding rate of time preference. However, it is a relatively simple matter to deal with more complex patterns of returns.

We capture most of the interesting aspects of this generalization by supposing that our typical government project can be described by two time sequences of numbers: b_t denoting the current dollar (consumption equivalent) value of the government service provided in period t, and e_t denoting the additional resources which must be raised in period t to carry out the project. Note that in any given period either b_t or e_t or both *may be negative.* A negative value of b_t corresponds to costs such as environmental damage caused by the project. A negative value of e_t may result when the project yields a cash return—for example, its output is sold on the market—which may be applied toward a reduction in taxes in that period. Negative values for these variables may occur in another way as well. Our analysis applies equally to investment and disinvestment. *Not* undertaking a project is itself a project; if project A is described by $\{b_t, e_t\}$, the project "not undertaking A" is described by $\{-b_t, -e_t\}$.[5]

Things will be simplified somewhat if we define shadow prices for benefits and revenues as follows:

$$\beta_t = 1 + \sum_{j=1}^{m} \alpha_t^j (v_t^j - 1) \tag{14}$$

$$\epsilon_t = 1 + \sum_{j=1}^{m} a_t^j (v_t^j - 1)$$

where the sums in (14) are taken over all of the private sector investment possibilities distinguished by the model. The term β_t is the social value (in dollars of time t consumption) of a dollar's worth of benefits when account is taken of the induced effects on private capital formation. Similarly, ϵ_t is the social value of the lost consumption and capital formation resulting from raising an additional dollar of finance.

In period t the government project produces b_t dollars in benefits, worth $b_t\beta_t$ dollars in period t consumption when the influence on private capital formation is taken into account. The project requires that e_t dollars in revenue be raised at a cost equivalent to $e_t\epsilon_t$ dollars of period t consumption when the influence on private capital formation is taken into account. For convenience let the point of reference be called t=0. The net effect is $b_t\beta_t - e_t\epsilon_t$ which is equivalent to $\delta_t(b_t\beta_t - e_t\epsilon_t)$ units of period 0 consumption, where δ_t is the *discount factor* derived from the social time preference term structure in the familiar manner. The project is worthwhile if

$$\sum_{t=0}^{T} \delta_t(b_t\beta_t - e_t\epsilon_t) \geq 0 \tag{15}$$

where the project has its last direct payoff or resource requirement in period T. It is easily checked that condition (15) gives us back the criterion for one-period investments (12) for which $b_{t+1}=1$, $b_\tau=0$ otherwise; $e_t=1$, $e_\tau=0$ otherwise.

---

[5] Note that we are here assuming only one form of benefits and only one form of finance. More generally, we could describe a project by a sequence of vectors $(b_t^1, b_t^2, \ldots, b_t^s, e_t^1, \ldots, e_t^r)$, where b_t^j = dollars worth of benefits of the jth type (of s types in all) in period t, and e_t^k = dollars required from finance source k (of r sources, e.g., sales tax, debt issue, etc.) in period t.

For the case in which an additional dollar of benefits is divided among consumption and the different sorts of private investment in the same way every year and the shadow prices of the different sorts of private capital are constant, $\beta_t=\beta$, a constant. If an additional dollar of financing comes out of consumption and the different sorts of investment in the same proportions every year and if the shadow prices of capital are constant, then $\epsilon_t=\epsilon$, a constant. Then (15) becomes

$$\sum_{t=0}^{T} \delta_t \left(b_t - \frac{\epsilon}{\beta} e_t\right) \geq 0 \tag{16}$$

investments should be evaluated by discounting at the social time preference rate, with government cash flows inflated by a factor ϵ/β. For the one-sector case we have seen a plausible *extreme* range for this weighting factor is $1/v \leq \epsilon/\beta \leq v$, with $\epsilon/\beta=1/v$ for $a_t=0$, $\alpha_t=1$, and $\epsilon/\beta=v$ for $a_t=1$, $\alpha_t=0$. When $\alpha_t=a_t=$ constant (changes in financing have the same impact on private investment as changes in value of government services), $\epsilon/\beta=1$, the criterion becomes simply to discount all flows, unweighted, at the social rate of time preference.

V. Concluding Remarks

In a sense, condition (15) with its associated definitions is the general solution to the public investment problem and hence the general conclusion of this analysis. The next step required is to put empirical flesh on the theoretical structure in the form of actual estimates of the various coefficients of condition (15). However, for various reasons, one should not be overoptimistic about obtaining a set of coefficients which can be applied to any government investment problem to lead to a correct choice.

First, the precise financing technique providing the source of funds for any government project is often unclear or not even well defined, and these sources vary from project to project. Do the funds to pay for a subsidy to the merchant marine come from corporation income taxes? Personal income taxes? Changes in federal debt? A correct answer would not even be assured if expenditure laws specified the source of finance. For example, federal highway programs are nominally financed by gasoline taxes, and this may be the end of the story. However, it is also possible that gasoline taxes would be about what they are anyway. An increase in highway expenditures in this case forces some other program to find its financing in another revenue device, say, the personal income tax. Then the source of finance for the highway expenditure for our purposes is in fact the personal income tax. Furthermore, the source of finance for a project may well be funds which would otherwise have financed another government activity. In this case, then, the dollars for the project in question "cost" the foregone benefits from the alternative activity. We need not labor further the extreme difficulty of establishing the financing source.

Second, knowing the financing source, we must establish how much of an incremental dollar from that source derives from consumption, how much from various forms of private capital formation (and, we might add, how much from unemployed resources). It is not sufficient to know simply the nominal payers of a tax, as it is the ultimate payers whose savings are affected.

Third, on the output side, matters are not much better. Assuming, as we have been, that one can reasonably estimate dollar values of the services from the government project, determining the influence of these flows on private investment is likely to be extremely difficult. Some headway may be made by examining the character of the flow involved. Is it more like consumption or more like savings? If the government undertakes to provide medical care for the aged, presumably this reduces the incentive for citizens to accumulate a reserve against this possibility. We might expect, then, that providing an additional dollar's worth of this insurance protection will lead individuals to reduce their private savings by about a dollar and to increase consumption by a like amount. At the other extreme, benefits from public parks may be effectively pure substitutes for private consumption. A family receiving these benefits reduces its expenditures on film and baseball game attendance, and increases its savings accordingly.

Note, though, that what a service looks like may tell only part of the story. The park example illustrates this. While it is true that the recreational services are of a nature which we would usually label "consumption," there is no guarantee that these services replace other consumption. They may be additive; they may even induce a reduction in labor supply and a net decrease in private saving ($\alpha<0$).

In view of these very serious obstacles to a precise implementation of (15), we would do well to consider what rules of thumb are likely to make sense and the circumstances in which they are likely to lead us astray. The argument in this paper suggests the simple principle: *Maximize present value of net dollar flows (including dollar equivalents of nonmarketed effects), discounting at the social rate of time preference.* As we have seen, assuming v_t, α_t, and a_t constants, reasonable extremes of condition (15) are given by the cases of (16) corresponding to $\beta=v$, $\epsilon=1$ and $\beta=1$, $\epsilon=v$. In the latter case, in which all financing comes out of investment and all benefits increase consumption, we should multiply financing changes by a factor v before calculating net benefits. If we nevertheless use our rule of thumb we shall, of course, undertake some socially unprofitable projects. For a project just barely worthwhile according to the rule so that $\Sigma\delta_t b_t=\Sigma\delta_t e_t$, undertaking the project will lead to a net loss equal to $(v-1)\ \Sigma\delta_t e_t$. At the other extreme, where all yield from the project leads to an equivalent value increase in private investment and all finance for the project derives from consumption ($\beta=v$, $\epsilon=1$), the rule of thumb is too conservative. The apparently barely worthwhile project will actually generate a net profit of $(v-1) \cdot\Sigma\delta_t e_t$.

Just how serious are these extreme errors? They are obviously sensitive to the value of v, the shadow price of private capital. This is an empirical matter but, forced to make a guess, and looking at Table 2, I would think $1.10 to be on the high side and $1.05 to be reasonable. If the *extreme* risks of error run the range from -5 percent to $+5$ percent of the value of the resources involved, modest in comparison, say, with the acceptable range of error of cost estimates, then the simplification effected by the rule of thumb may be well worth the price. Even if this is accepted, of course, there remains the critical issue of the social choice of time preference rate, since the particular number chosen is of greatest consequence for decisions among projects of any considerable duration. The choice of time preference rate, essentially involving questions of distribution of welfare among generations, leads us into the literature on optimal savings and away from the subject of this paper.

REFERENCES

K. J. Arrow, "Discounting and Public Investment Criteria," in A. V. Kneese and S. C. Smith, eds., *Water Resources Research,* Baltimore 1966.

——— and M. Kurz, *Public Investment, The*

Rate of Return and Optimal Fiscal Policy, Baltimore 1970.

W. J. Baumol, "On the Social Rate of Discount," *Amer. Econ. Rev.,* Sept. 1968, *58,* 788–802.

D. F. Bradford, "Constraints on Public Action and Rules for Social Decision," *Amer. Econ. Rev.,* Sept. 1970, *60,* 642–54.

P. Diamond, "The Opportunity Costs of Public Investment: Comment," *Quart. J. Econ.,* Nov. 1968, *82,* 682–88.

M. S. Feldstein, "Financing in the Evaluation of Public Expenditure," disc. pap. no. 132, Harvard Inst. Econ. Res., Aug. 1970.

A. Sandmo and J. H. Drèze, "Discount Rates for Public Investment in Closed and Open Economies," *Economica,* Nov. 1971, *38,* 395–412.

The Demand for Risky Assets

By IRWIN FRIEND AND MARSHALL E. BLUME*

A striking characteristic of recent research on capital asset pricing is the paucity of empirical work on the determinants of the market price of risk[1] in contrast to the abundance of work on the interrelationships of the risk premiums among different risky assets. The market price of risk will in part depend on the utility functions of individual investors and, except for a special class of functions, on the distribution of wealth among investors.

The relationship between utility functions and wealth of all investors is obviously basic to constructing an aggregate demand function for risky assets. However, there are a multiplicity of other important potential uses, including the derivation of the form of household consumption or saving functions, measurement of the impact of different government, fiscal, or insurance measures on economic welfare, and assessment of the problems in extending single period financial and economic investment decisions to a multiperiod world.

Until recently, most interested economists have seemed to believe that utility functions of individuals were characterized either by constant or decreasing absolute risk aversion and by increasing proportional risk aversion (Section I). The measure of proportional risk aversion used in this paper is the same as the elasticity of the marginal utility of wealth. In recent years, this traditional view of increasing proportional risk aversion has been vigorously questioned by advocates of the *log* utility function which, while still maintaining decreasing absolute risk aversion, implies that proportional risk aversion is constant rather than increasing.

This paper will, for the first time, systematically exploit cross-sectional data on household asset holdings to assess the nature of households' utility functions. The data used are detailed information on assets, income, and other socioeconomic-demographic characteristics for a large sample of households.

Prior to analyzing the survey data, the paper will adapt and extend existing theory to obtain the relationships between the composition of household wealth (both human and nonhuman) and their utility functions. These relationships are suitable for statistical analyses at both the micro and macro levels (Section II). Part of the empirical work is based upon these survey data (Section III). Other work analyzes *ex post* and *ex ante* market returns since the latter part of the nineteenth century to estimate the market price of risk (Section IV). The last part of the paper synthesizes the analyses of the household and market data (Section V).

Our main conclusions are: First, regardless of their wealth level, the coefficients of proportional risk aversion for households are on average well in excess of one and probably in excess of two. Thus, investors require a substantially larger premium to

---

* Professor of finance and economics, and professor of finance, respectively, University of Pennsylvania. We wish to thank the National Science Foundation and the Rodney L. White Center for Financial Research for financial support.

[1] The market price of risk is the difference between the expected rates of return on the market portfolio of risky assets as a whole and on a risk-free asset per unit of risk of the market portfolio. For purposes of this paper, it turns out that the relevant measure of risk is the variance of returns.

hold equities or other risky assets than they would if their attitudes toward risk were described by logarithmic utility functions.

Second, the paper concludes that the assumption of constant proportional risk aversion for households is as a first approximation a fairly accurate description of the market place. These first two findings imply that the utility function of a representative investor is quite different from those which have often been assumed in the literature. However, it should be pointed out that our conclusion of constant proportional risk aversion follows from our treatment of investment in housing. Other plausible treatments would imply either moderately increasing or moderately decreasing proportional risk aversion.

Third, under tenable assumptions, including that of constant proportional risk aversion, we develop a simple form of the aggregate equilibrium relationship between the relative demand for risky assets and the market price of risk. To determine the actual values of the required rate of return on risky assets, the market price of risk, and the relative value of risky assets would of course entail the specification of supply conditions. For a constant physical supply of both risky and nonrisky assets, the relationship developed in this paper indicates how the required rate of return on risky assets as a whole is determined.

I. Earlier Studies

Economists have generally been convinced that the market utility function has risk-aversion properties somewhere between a negative exponential utility function with constant absolute risk aversion and increasing relative risk aversion, and a constant elasticity utility function with decreasing absolute risk aversion, but with constant proportional risk aversion. The authority generally cited for asserting these bounds for the utility function is Kenneth Arrow (for example, see S. C. Tsiang).

While no one is likely to argue with the plausibility of decreasing absolute risk aversion, the widely held assumption of increasing (or at most constant) relative risk aversion is open to question. At the theoretical level, the denial of the tenability of decreasing relative risk aversion is based on the assumed implausibility of a utility function which is unbounded either from above or from below (see Arrow). However, arguments based upon such bounding conditions may be economically empty. The ultimate justification for such an assumption must rest on the empirical data. The only direct evidence we know of that has been cited to support the assumption of increasing or constant relative risk aversion are the studies which conclude that either the income elasticity or the wealth elasticity of demand for cash balances (usually money narrowly or broadly defined) is at least one (see Arrow). Of this evidence, only the wealth elasticity is at all relevant. Moreover, in view of the existence of liabilities or negative assets in investors' balance sheets, it is more appropriate to relate the total of risky assets, rather than the customary cash balances, to net worth, but this does not appear to have been done in the relevant literature.

The available cross-section studies based on household surveys of assets and liabilities seem to point to a wealth elasticity of liquid assets of well below or close to one if all tangible assets including consumer durables are included in wealth, and lower figures if tangible assets are excluded (for example, see Jean Crockett and Friend, pp. 37 and 55–57). The aggregate time-series data reflecting changes in supply and demand conditions, including those arising from changes in wealth distribution among different groups in the population, appear to be much less pertinent than the cross-section information.[2]

In recent years, the major challenge to the widely held assumption of a market utility function characterized by increasing relative risk aversion has been posed by proponents of the *log* utility function which implies constant relative risk aversion (again associated with decreasing absolute risk aversion). The *log* utility function has been justified in this literature mainly on grounds of the reasonableness of the implicit growth maximization criterion as contrasted with the alleged unreasonableness of mean-variance efficient portfolios.[3]

More recently, in Richard Roll's paper a new set of empirical tests of the implications of the growth-optimum and mean-variance models found that the former performs well in comparison with the latter "but the test results are clouded by the close operational similarity of the two models" (p. 552). Still another set of empirical tests obtained from experimental games suggested that a *log* utility function might characterize individuals with wealth in excess of $200,000, but below that level increasing relative (and decreasing absolute) risk aversion was exhibited. (See M. J. Gordon, G. E. Paradis, and C. H. Rorke.)

II. Theoretical Background for the Empirical Analysis

The literature has shown under various conditions that in equilibrium the market price of risk—the ratio of the expected risk premium on risky assets to the variance of return on those assets—equals the market value of all risky assets times a function of each individual's measure of absolute risk aversion. More precisely, this relationship can be developed for discrete planning periods by assuming either that investors' utility functions are quadratic or that their utility functions are exponential or logarithmic and end-of-period wealth is normally distributed.[4] This relationship takes the form

$$(1) \qquad \frac{E(r_m - r_f)}{\sigma^2(r_m)} = V_{mo}\left[\sum_k E\left(\frac{1}{R_k}\right)\right]^{-1}$$

where r_m is the return on the market portfolio of all risky assets, r_f is the return on the risk-free asset, R_k is Pratt's measure of absolute risk aversion for the kth investor, and V_{mo} is the market value of all risky assets at the beginning of the planning period. Stephen Ross has obtained a similar expression by assuming an infinitesimal planning horizon and no finite changes in value of any asset in an infinitesimal period.[5]

Although the various developments of (1) assumed that the net supply of the risk-free asset was zero, they can easily be generalized to the situation in which this asset has a positive net supply. For mathematical convenience, this generalization will be based upon an infinitesmal horizon model and will make the same assumptions as those made by Ross. Besides the assumption about the distribution of returns mentioned above, perhaps the most critical assumptions include homogeneity of expectations and a frictionless capital market. In a frictionless capital market, all financial assets—both risk-free and risky—are infinitely divisible, can be traded with no transaction costs, and can be sold short with the proceeds available to purchase another asset.

---

[2] Even the time-series analyses are not consistent in indicating a wealth elasticity of cash balances, broadly defined, equal to or greater than one in the period following World War II. See Allan Meltzer, p. 236.

[3] See Henry A. Latané; Nils Hakansson; and Harry Markowitz. For a different theoretical position, see Paul Samuelson and Robert Merton.

[4] See Jan Mossin; John Lintner; and A. P. Budd and R. H. Litzenberger.

[5] This assumption is nontrivial and would exclude such distributions of returns of individual assets as nonnormal stable.

In parallel with Ross's development, the wealth conservation equation for the kth investor, assuming no taxes, would be[6]

$$(2)\quad W_{k,t+dt} = W_{kt}\{1 + [r_f + \alpha_k E(r_m - r_f)]dt + \alpha_k \sigma_m y(t)\sqrt{dt}\}$$

where α_k is the proportion of the net worth of investor k placed in the portfolio of risky assets, σ_m is the standard deviation of the returns on the portfolio of risky assets, and $y(t)$ is a standardized normal random variate.[7] By expanding $U(W_{k,t+dt})$ about W_{kt}, taking expected values and dropping terms involving dt to the power of 2 or more, one obtains

$$(3)\quad E[U(W_{k,t+dt})] = U(W_{kt}) + U'(W_{kt})W_{kt}[r_f + \alpha_k E(r_m - r_f)]dt + \tfrac{1}{2}U''(W_{kt})W_{kt}^2\alpha_k^2\sigma_m^2 dt$$

To determine the optimal value of α_k, the derivative of $E[U]$ is set to zero:

$$(4)\quad U'(W_{kt})E(r_m - r_f) + U''(W_{kt})W_{kt}\alpha_k\sigma_m^2 = 0$$

Recalling that Pratt's measure of relative risk aversion C_k is defined as $W_{kt}[-U''(W_{kt})/U'(W_{kt})]$,[8] (4) can be rewritten as[9]

$$(5)\quad \alpha_k = \frac{E(r_m - r_f)}{\sigma_m^2}\cdot\frac{1}{C_k}$$

In the absence of taxes, (5) could be used to estimate the coefficient of relative risk aversion for investor k given estimates of α_k and the market price of risk. However, if returns from both risky and risk-free assets are taxed at the same rates in any tax bracket[10] and if t_k is the average rate of tax for investor k, equation (2) could be modified to incorporate the tax effect. Repeating the logic underlying (2) through (5) would yield the tax adjusted relationship:

$$(6)\quad \alpha_k = \frac{E(r_m - r_f)}{\sigma_m^2}\cdot\frac{1}{(1 - t_k)C_k}$$

Since r_m and r_f are measured before personal taxes, the assumption of homogeneous expectations assures that if $(1-t_k)\alpha_k$ is constant over all investors, C_k will also be constant. This observation will be used in the next section in conjunction with cross-sectional survey data to assess how C_k varies with net worth.

So far, it has been explicitly assumed that all assets are liquid in that they can be purchased or sold at no cost in any quantity. This assumption is clearly unrealistic for human wealth. David Mayers has shown in a mean-variance world of finite horizons how this assumption can be relaxed to allow for the nonmarketability of human wealth. His insight can be readily applied to the continuous case. To do this, redefine r_m as the return on the portfolio of liquid risky assets and α_k as the proportion of investor k's liquid net worth (total net worth less human wealth) placed in the portfolio of liquid risky assets. Further, define the new terms: h_k as the ratio of the value of the human wealth of the kth investor to his net worth and r_{hk} as the rate of return on the kth investor's human wealth.

---

[6] An intuitive explanation for the appearance of the standard deviation with $\sqrt{dt}$ rather than with dt is that if a time period is subdivided into n subperiods each of duration $\tau = 1/n$, and if the returns over each of these subperiods are independently distributed, then the variance of returns over τ, σ_τ^2 is equal to $1/n$ times the variance of returns σ_t^2 over the initial period. As n approaches ∞, $1/n$ approaches dt, so that σ_τ approaches $\sigma_t\sqrt{dt}$.

[7] Equation (2) implies that the kth individual invests α_k of his assets in risky assets and $1-\alpha_k$ in the risk-free asset. In the type of continuous time model used in this paper, the separation theorem holds.

[8] This measure of relative risk aversion can also be interpreted as the wealth elasticity of the marginal utility of wealth as indicated in the introduction.

[9] Samuelson has obtained a similar expression for the individual investor.

[10] This assumption is tantamount to assuming that capital gains and ordinary income are taxed identically and abstracts from the special tax treatment of homes and tax-exempt securities. However, see fn. 37.

Under these definitions, the wealth conservation equation adjusted for taxes t_k, again assumed the same for all types of income, would be

$$(7)\quad W_{k,t+dt} = W_{kt}\Big\{1 + (1-t_k) \cdot \{(1-h_k)[r_f + \alpha_k E(r_m - r_f)] + h_k E(r_{hk})\}dt + (1-t_k) \cdot [(1-h_k)^2\alpha_k^2\sigma_m^2 + h_k^2\sigma_{hk}^2 + 2\alpha_k h_k(1-h_k) \cdot Cov\,(r_m, r_{hk})]^{1/2} y(t)\sqrt{dt}\Big\}$$

The covariance term allows for the presence of dependencies between the returns on liquid and human wealth. That both the variance of human wealth and the covariances are subscripted by k means that these statistics can vary from one individual to another. Repeating the logic underlying (2) through (5) would yield

$$(8)\quad \alpha_k = \frac{E(r_m - r_f)}{\sigma_m^2} \cdot \frac{1}{(1-t_k)(1-h_k)C_k} - \frac{h_k}{1-h_k}\beta_{hk,m}$$

where $\beta_{hk,m}$ is the ratio of $Cov(r_m, r_{hk})$ to σ_m^2. Thus, $\beta_{hk,m}$ can be interpreted as the slope coefficient of the regression of r_{hk} on r_m. It is interesting to note that the ratio α_k does not depend upon $E(r_{hk})$ but only upon the dependence between r_m and r_{hk}.[11]

Equations (6) and (8) are micro functions for the individual investor and are not the same kinds of relationships as (1), which is an equilibrium condition for the market as a whole—a macro-economic concept. Such macro relationships can be developed from these micro functions by forming a weighted average of α_k over all investors, the weights defined as the ratio of the investor's liquid wealth to the total liquid wealth in the market. If γ_k equals this proportion, if α is defined as the ratio of liquid risky assets to all liquid assets, and if all assets are liquid, the micro relationship (6) would imply

$$(9)\quad \sum \gamma_k\alpha_k = \alpha = \frac{E(r_m - r_f)}{\sigma_m^2}\sum \frac{\gamma_k}{(1-t_k)C_k}$$

Relation (9), which is easily solved for the market price of risk, differs from (1) mainly in that it takes into account taxes, allows the net supply of the risk-free asset to be positive, and is expressed in terms of relative rather than absolute risk aversion.

If one includes human wealth and further assumes that this wealth is illiquid, a similar macro relationship to (9) can be derived by averaging α_k given by (8). The resulting average is

$$(10)\quad \sum \gamma_k\alpha_k = \alpha = \frac{E(r_m - r_f)}{\sigma_m^2} \cdot \left[\sum \frac{\gamma_k}{(1-t_k)(1-h_k)C_k}\right] - \sum \gamma_k \frac{h_k}{1-h_k}\beta_{hk,m}$$

The macro relationship (10), again easily solved for the market price of risk, is quite different from (1) and (9) in that it takes explicit account of the nonmarketability of human wealth.

---

[11] Since the dependence between r_m and r_{hk} is a factor in determining the optimum ratio of risky liquid assets to all liquid assets, it would be anticipated that an investor might be able to improve his expected utility by altering the proportions of the risky liquid assets he holds from those implicit in the market portfolio to take into account differences in the values of the covariances of human wealth and individual liquid assets in the market portfolio. The work of Y. Landskroner indeed shows this to be the case. Whether (8) results in an expected utility close to the optimum depends upon the correlation of the optimal portfolio of risky liquid assets and the market portfolio. If this correlation were very close to 1.0, for instance, it would be possible to construct a portfolio of the market and the risk-free asset which would have a probability distribution very similar to the optimal portfolio of liquid assets. Numerous empirical studies have demonstrated that the returns on portfolios of a large number of securities are highly correlated with the market portfolio.

It should be noted that equations (9) and (10), and the simplifications of these equations presented subsequently, represent aggregate equilibrium relationships between the relative demand for risky assets and the market price of risk. The market price of risk at the macro level is not given, as it may be at the micro level, but is determined jointly with the aggregate ratio α of liquid risky assets to all liquid assets. For example, assuming a constant physical supply of both risky and nonrisky assets, α is determined by $E(r_m)$ and r_f. Equations (9) and (10) indicate how $E(r_m)$ and α are related to C_k, the distribution of wealth γ_k, tax rates t_k, σ_m^2, and r_f.[12]

If an investor's coefficient of relative risk aversion is assumed independent of γ_k and this coefficient is also independent of his average tax rate, (9) can be vastly simplified. This assumption of constant proportional risk aversion on the average is considerably more general than the assumption that all investors have the same value for C_k in that C_k can be a function of characteristics of investors which are independent of γ_k and t_k.

Specifically, let it be assumed that C_k is given by

$$\frac{1}{C_k} = \frac{1}{C} + \epsilon_{ck} \tag{11}$$

where ϵ_{ck} is a mean-zero disturbance independent of γ_k and t_k. The symbol C defined by (11) can be interpreted as the harmonic mean of C_k. Let us first simplify (9) by substituting (11) into it to obtain

$$\alpha = \frac{E(r_m - r_f)}{\sigma_m^2} \cdot \left[\frac{1}{C}\sum \frac{\gamma_k}{1 - t_k} + \sum \frac{\gamma_k \epsilon_{ck}}{1 - t_k}\right] \tag{12}$$

The application of expected value operators to (12) causes the second summation in the brackets to go to zero. Since the sum of the γ_ks is one, the reciprocal of the first summation might be termed the weighted harmonic mean of $(1-t_k)$. Designating this mean as $1-t$ and solving for the market price of risk, (12) takes the simplified form

$$\frac{E(r_m - r_f)}{\sigma_m^2} = \alpha C(1 - t) \cdot \tag{13}$$

If in addition to the assumption of constant proportional risk aversion (on the average) it can further be assumed that $\beta_{hk,m}$ and ϵ_{ck} are independent of the level of human wealth of the kth individual, (10) can likewise be vastly simplified. To do this, note first that the terms α, $(\gamma_k h_k)/(1-h_k)$, and $\gamma_k/(1-h_k)$ can be rewritten as R/L, H_k/L, and W_k/L, respectively, where H_k is the value of the kth investor's human wealth, W_k his net worth, R the total value of all risky liquid assets, and L the total value of all liquid assets. With these identities, the simplification of (10) proceeds by substituting (11) into (10), taking expected values, substituting these identities, and solving for the market price of risk. The resulting expression is

$$\frac{E(r_m - r_f)}{\sigma_m^2} = \left[\frac{R}{W^*} + \beta_{hm}\frac{H}{W^*}\right] C \tag{14}$$

where W^* is $\sum W_k/(1-t_k)$ or a tax adjusted sum of all wealth and β_{hm} is defined as $E(\beta_{hk,m} | H_k)$, assumed the same for all investors.[13]

---

[12] This can easily be seen by writing in a one-period model

$$\alpha = \frac{E(R_m)}{1 + E(r_m)} \div \left[\frac{E(R_m)}{1 + E(r_m)} + \frac{F}{1 + r_f}\right]$$

where $E(R_m)$ represents the expected dollar return over the period on the available supply of liquid risky assets at the beginning of the period, and F represents the corresponding known dollar return on riskless assets. The constant physical supply of liquid risky assets and risk-free assets are measured by $E(R_m)$ and F, respectively.

[13] Though (14) was developed under the restrictive assumption that the sole decision of the individual was how much of his liquid assets to place in the market

(over)

These simple equilibrium expressions are valid if investors' utility functions are characterized by constant proportional risk aversion on the average as specified in (11). If valid, (13) and (14) would seem to offer tractable ways to estimate C from time-series data.[14] They also provide extremely simple forms of the aggregate equilibrium relationship between the relative demand for risky assets and the market price of risk.

III. The Analysis of the Survey Data

The micro relationships just developed lend themselves to the use of cross-sectional data in ascertaining whether or not C_k, the coefficient of proportional risk aversion, is invariant on the average to investors' net worth. After describing the cross-sectional data, we will first examine the behavior of C_k employing (6) which implicitly assumes that all wealth is liquid and then will assess it employing (8) which explicitly accounts for the nonmarketability of human capital.

The cross-sectional data to be analyzed come from the 1962 and 1963 Federal Reserve Board (*FRB*) Surveys of the Financial Characteristics of Consumers and Changes in Family Finances. These surveys, which oversampled the upper income groups, collected for more than 2,100 households detailed information on the value of their assets and liabilities at the end of both 1962 and 1963 and on the sources and amounts of income in both of these years.

From these data, the study constructed three different types of balance sheets at the end of 1962. The first type included all assets and associated liabilities with the exception of human wealth and homes.[15] The second excluded only human wealth. The third included not only homes, but also an estimate of human wealth. These balance sheets, expressed as ratios to the corresponding measures of net worth, are summarized in Tables 1–3 by net worth categories. Changes in the definition of wealth will, of course, move households from one wealth class to another.

In checking the validity and reasonableness of the data, it was observed that a few households received substantial salaries from closely held businesses in which they had active interests. Since some part of these salaries might more properly be classified as dividends or return on capital, any salaries from such closely held businesses in excess of $25,000 were valued as perpetuities at 10 percent and added to equity. To determine how sensitive the results are to this adjustment, the subsequent analyses were replicated on a subsample which excludes any households which own such businesses. The conclusions were virtually unchanged.[16]

Only households with a net worth in excess of $1,000 were included. Since such households have a predominant impact upon the market for assets, this restriction would have little effect upon the value of the variables in the macro relationships

---

portfolio of liquid risky assets, Landskroner has shown that if r_0, the return on an asset uncorrelated with the return on the market portfolio of risky assets, is substituted for r_f and is also uncorrelated with the return on human capital, (14) holds in the more general case where an investor's decision variables include not only how much to put into risky liquid assets, but also how much to place into each risky liquid asset.

[14] To utilize (13) in a time-series, the weighted harmonic mean of $(1-t_k)$ might be approximated by the ratio of total disposable income to total adjusted gross income. For the survey of households to be examined in the next section, the weighted harmonic mean would be 0.787 while the ratio of total disposable income to total adjusted gross income would be 0.859 which is close to the 0.871 implied by the *Statistics of Income*. Thus, if this harmonic mean is representative of the population, the ratio of total disposable income to total adjusted gross income would be about 10 percent greater than the harmonic mean. This figure of 10 percent might be useful in adjusting the time-series available from the *Statistics of Income* to that required by (13).

[15] With homes are included automobiles which represent only a minor proportion of assets except possibly for the lowest wealth groups.

[16] See fn. 21.

Table 1—Average Ratios of Assets and Selected Items to Household Net Worth Exclusive of Homes (and Associated Mortgages) for Households Classified by Net Worth[a]

December 31, 1962

	Net Worth (Thousands of Dollars)					
Type of Item	1–10	10–100	100–200	200–500	500–1,000	Over 1,000
Risk-Free Assets to Net Worth						
Checking Accounts	0.079	0.026	0.020	0.014	0.018	0.009
Other Cash Balances[b]	0.301	0.192	0.062	0.029	0.027	0.021
Savings Bonds	0.069	0.040	0.034	0.012	0.012	0.003
Life Insurance (Cash Value)	0.238	0.071	0.042	0.025	0.031	0.013
Other Risk-Free Assets[c]	0.127	0.053	0.004	0.010	0.017	0.012
	0.814	**0.382**	**0.163**	**0.089**	**0.103**	**0.057**
Mixed-Risk Assets to Net Worth						
State and Local Bonds	0.0	0.001	0.001	0.013	0.014	0.032
Other Mixed-Risk Assets[d]	0.042	0.077	0.131	0.084	0.064	0.099
	0.042	**0.078**	**0.132**	**0.098**	**0.078**	**0.131**
Risky Assets to Net Worth						
Common and Preferred Stock	0.050	0.111	0.272	0.363	0.283	0.316
Equity in Uninc. Business	0.081	0.308	0.286	0.370	0.315	0.328
Other Risky Assets[e]	0.140	0.167	0.186	0.104	0.265	0.225
	0.271	**0.586**	**0.743**	**0.838**	**0.863**	**0.869**
Total Assets to Net Worth	**1.128**	**1.046**	**1.039**	**1.025**	**1.004**	**1.057**
Liabilities to Net Worth	0.128	0.046	0.039	0.025	0.044	0.057
Ancillary Statistics						
$(1-t_k)\alpha_k$[f]	0.285	0.601	0.749	0.763	0.677	0.650
Number of Households	523.	477.	68.	100.	69.	103.

[a] These averages are weighted by the inverse of the sampling probability for each household. These sampling probabilities have been adjusted for nonresponse and other factors. The net worth used in classifying households is defined in the same terms as the balance sheet.

[b] Includes checking and other commercial bank accounts, savings and loan savings accounts, credit union savings accounts, and mutual savings accounts.

[c] Includes U.S. Treasury Bills, notes and certificates, the withdrawal value of profit sharing and retirement plans, credit balances in brokerage accounts, and risk-free assets held in trust accounts. Risky assets held in trust accounts are included in the appropriate category of direct holdings, except that when they could not be classified in this manner, they were included in miscellaneous risky assets.

[d] Includes long-term corporate, state and local and U.S. government bonds (other than savings bonds).

[e] Includes investment real estate assets and miscellaneous assets, such as patents, etc.

[f] The symbol α_k is defined as the ratio of the sum of the mixed-risk and risky assets to net worth for investor k. The symbol t_k is the average federal tax rate for investor k as estimated by the procedure devised by the Survey of Financial Characteristics of Consumers. The only difference was that realized capital gain income or losses were included in adjusted gross income in this paper unlike the original procedure. The original survey ignored this type of income in calculating the tax rate.

(13) or (14). Yet, it appears to improve the quality of the survey data by eliminating those households who may have underreported their assets or whose assets were temporarily lower than normal.

A crude estimate of human wealth was given by the discounted value of the average labor income in 1962 and 1963 which was then assumed to grow at 4 percent per year. The discount rate was taken to be 10 percent. If a person was less than 65, he was assumed to retire at 65. If he was working and over 65, he was assumed to continue to work for four years; if over 69, three years; if over 74, two years; and if over 79, one year. The sensitivity of the subsequent results to this arbitrary method of calculating human wealth will be examined, and it will turn out that within broad ranges of discount rates and growth rates

TABLE 2—AVERAGE RATIOS OF ASSETS AND SELECTED ITEMS TO HOUSEHOLD NET WORTH INCLUSIVE OF HOMES FOR HOUSEHOLDS CLASSIFIED BY NET WORTH[a]

December 31, 1962

	Net Worth (Thousands of Dollars)					
Type of Item	1–10	10–100	100–200	200–500	500–1,000	Over 1,000
Risk-Free Assets to Net Worth						
Checking Accounts	0.040	0.019	0.017	0.012	0.023	0.009
Other Cash Balances	0.138	0.109	0.084	0.032	0.023	0.019
Savings Bonds	0.029	0.025	0.020	0.020	0.009	0.003
Life Insurance (Cash Value)	0.097	0.056	0.033	0.025	0.033	0.013
Other Risk-Free Assets	0.055	0.033	0.013	0.008	0.010	0.012
	0.359	**0.243**	**0.167**	**0.096**	**0.097**	**0.056**
Mixed-Risk Assets to Net Worth						
State and Local Bonds	0.0	0.000	0.001	0.005	0.037	0.029
Other Mixed-Risk Assets	0.016	0.038	0.084	0.077	0.043	0.090
	0.016	**0.039**	**0.085**	**0.082**	**0.081**	**0.119**
Risky Assets to Net Worth						
Market Value of Homes	1.539	0.657	0.216	0.098	0.112	0.046
Common and Preferred Stock	0.012	0.052	0.208	0.289	0.306	0.292
Equity in Uninc. Business	0.045	0.147	0.271	0.395	0.235	0.339
Other Risky Assets	0.035	0.101	0.108	0.084	0.213	0.216
	1.631	**0.957**	**0.803**	**0.866**	**0.866**	**0.892**
Total Assets to Net Worth	**2.006**	**1.238**	**1.055**	**1.044**	**1.043**	**1.067**
Liabilities to Net Worth						
Mortgage Liabilities	0.924	0.204	0.041	0.016	0.008	0.003
Other Liabilities	0.083	0.034	0.014	0.028	0.035	0.064
	1.006	**0.238**	**0.055**	**0.044**	**0.043**	**0.067**
Ancillary Statistics						
$(1-t_k)\alpha_k$ (Equity in Homes)[b]	0.669	0.718	0.724	0.780	0.642	0.665
$(1-t_k)\alpha_k$ (Investment in Homes)[b]	1.495	0.897	0.758	0.793	0.648	0.667
Number of Households	459.	731.	94.	103.	76.	108.

[a] See fn. for Table 1.

[b] Includes automobiles.

the results are substantially unaffected.

If the micro relationship (6) were to hold, an estimate of the market price of risk times the reciprocal of the coefficient of relative risk aversion $[E(r_m-r_f)/\sigma_m^2]C_k^{-1}$ is given by $\alpha_k(1-t_k)$. The assumption made in the last section that all investors agree on the value of the market price of risk means that $\alpha_k(1-t_k)$ provides an estimate of C_k^{-1} up to a multiplicative positive constant, so that $\alpha_k(1-t_k)$ can be used to assess how C_k^{-1} and, thereby, how C_k varies with net worth.[17] Using either of the first two balance sheets, which do not include human capital, the cross-sectional data provide estimates of $\alpha_k(1-t_k)$. Since (6) does not take account of the nonmarketability of human wealth, the analysis based upon (6) will not include this kind of wealth. It might be mentioned that virtually all empirical applications of

---

[17] If $E[C_k^{-1}]$ were to decrease with increases in net worth W_k, it is not necessary that $E(C_k)$ would increase. Yet, if the distributions of C_k given W_k, $P(C_k|W_k)$, remain unchanged as W_k increases except for location, it can be shown that $E(C_k)$ would increase with increases in W_k. This proposition follows from noting that if for fixed ϵ_k, $[E(C_k)+\epsilon_k]^{-1}>[E(C_k)'+\epsilon_k]^{-1}>0$, $E(C_k)+\epsilon_k<E(C_k)'+\epsilon_k$. Since $P(C_k|W_k)$ is unchanged except for location, the distributions of ϵ_k will be the same at each level of W_k. Taking the definite integral of these two inequalities with respect to the distribution of ϵ_k yields the desired result.

TABLE 3—AVERAGE RATIOS OF ASSETS AND SELECTED ITEMS TO HOUSEHOLD NET WORTH INCLUSIVE OF HUMAN WEALTH AND HOMES CLASSIFIED BY NET WORTH[a]

December 31, 1962

	Net Worth (Thousands of Dollars)					
Type of Item	1–10	10–100	100–200	200–500	500–1,000	Over 1,000
Risk-Free Assets to Net Worth						
Checking Accounts	0.027	0.007	0.003	0.005	0.014	0.008
Other Cash Balances	0.116	0.042	0.014	0.021	0.019	0.016
Savings Bonds	0.011	0.011	0.003	0.008	0.005	0.003
Life Insurance (Cash Value)	0.027	0.013	0.010	0.014	0.021	0.013
Other Risk-Free Assets	0.000	0.004	0.006	0.013	0.009	0.012
	0.181	**0.077**	**0.036**	**0.061**	**0.068**	**0.052**
Mixed-Risk Assets to Net Worth						
State and Local Bonds	0.0	0.000	0.000	0.001	0.019	0.023
Other Mixed-Risk Assets	0.013	0.014	0.008	0.032	0.023	0.070
	0.013	**0.014**	**0.008**	**0.033**	**0.041**	**0.093**
Risky Assets to Net Worth						
Market Value of Homes	0.470	0.183	0.101	0.104	0.086	0.043
Common and Preferred Stock	0.002	0.017	0.018	0.085	0.171	0.253
Equity in Uninc. Business	0.055	0.070	0.022	0.085	0.220	0.277
Human Capital	0.334	0.656	0.850	0.649	0.349	0.148
Other Risky Assets	0.018	0.038	0.014	0.035	0.100	0.189
	0.878	**0.964**	**1.006**	**0.958**	**0.926**	**0.910**
Total Assets to Net Worth	**1.072**	**1.056**	**1.049**	**1.052**	**1.035**	**1.055**
Liabilities to Net Worth						
Mortgage Liabilities	0.052	0.044	0.043	0.039	0.013	0.004
Other Liabilities	0.021	0.012	0.006	0.013	0.022	0.051
	0.072	**0.056**	**0.049**	**0.052**	**0.035**	**0.055**
Ancillary Statistics						
Number of Households	114.	819.	485.	200.	90.	135.

[a] See fn. for Table 2.

portfolio theory have ignored human wealth in spite of its obvious importance to the demand for risky assets.

In interpreting the meaning of these numbers, it should be noted that the derivation of (6) assumed that all assets are acquired for investment purposes, can be unambiguously dichotomized as risky or risk-free, are subject to the same schedule of income taxes, and are infinitely divisible. For analyses of the composition of wealth inclusive of homes, these assumptions may be so unrealistic as to place serious restrictions on the validity of (6).

Contrary to the implicit treatment of assets in our model and in capital asset pricing theory generally, homes are typically acquired by households for consumption as well as investment purposes. Moreover, the lack of tax on the imputed income from owning a home as well as other subsidies make the tax rate assumption highly questionable. Furthermore, a house involves a substantial investment and for a person with limited liquid assets, the incurrence of more debt than he might normally want to have. In the absence of an adequate rental market, the purchase of a home may be the only way of obtaining the desired type of housing. Holding utility functions constant, these factors would tend to increase the values of α_k for the

Table 4—Regressions of $(1-t_k)\alpha_k$ on the Logarithm of Net Worth With and Without Holding Constant Age, Education, and Occupation

December 31, 1962

Net Worth Including Homes	Housing[a] Measured by	Simple Regression			Regression with Dummy Variables for Age, Education, and Occupation[b]		
		Coefficient on $ln(NW)$			Coefficient on $ln(NW)$		
		estimate	t-value	$\bar{R}^2$	estimate	t-value	$\bar{R}^2$
No		0.063	11.27	0.09	0.055	7.76	0.15
Yes	Equity	−0.006	− 1.52	0.00	−0.001	−0.23	0.05
Yes	Full Value	−0.185	−15.30	0.13	−0.155	−9.77	0.18

[a] Includes automobiles.

[b] The dummy variables, assuming the value of 1.0 if the characteristic is applicable to the head of the household and 0.0 otherwise, are for age (less than or equal to 25 and 5-year intervals from 25 to 65), for education (grade school or less, some high school, high school graduate, some college, or college graduate), and for occupation (self-employed, employed by others, retired, farm operator, not gainfully employed, and employer unknown).

lower wealth groups relative to those for the upper wealth groups if the investment in a home is assumed to be risky.[18]

Since investments in homes are not strictly compatible with the assumptions underlying (6), homes have been treated in three different ways in the empirical work: not as an investment asset; as a risky asset where the relevant investment is measured by the household's equity in the home; and as a risky asset where the investment is measured by the gross market value of the home. Another possibility would be to assume that households consider homes to be riskless assets, but this seems a less tenable assumption than the ones we have utilized.

Within the context of the model embodied in (6), perhaps the most appropriate measure of a household's investment in housing is its equity value. First, if an investor views a mortgage and a house as a package, it may well be that he regards only his equity at risk. Second, the use of equity is more in the spirit of infinitely divisible assets than the use of total assets. Nonetheless, the results of alternative assumptions are presented.

Tables 1 and 2 present averages of the estimated tax adjusted ratio $(1-t_k)\alpha_k$ by net worth classes for varying treatments of housing. Human wealth is not included. In addition, Table 4 presents the slope coefficients of the regressions of $(1-t_k)\alpha_k$ on the logarithm of net worth.[19] The α_k in these tables are ratios of risky and mixed-risk assets to net worth. Both simple regressions and regressions which control with dummy variables for the possible interaction between net worth and three socioeconomic factors (age, education, and occupation) are presented. Whether or not these dummy variables are included, the slope coefficients on the logarithm of net worth are not much different.[20]

---

[18] Housing presents other difficulties in applying the micro relationships developed in the last section. First, since an investor values a house not only for its monetary rewards but also for its consumption value, it may be difficult to aggregate returns from housing with those from other assets. Second, the purchase of a home with a large mortgage permits an investor to hedge against changes in housing costs and in the costs of borrowing funds.

[19] In the first attempt at running these regressions, the sampling probability associated with each household was used to adjust for heteroscedasticity on the assumption that the reciprocal of the sampling probabilities in each stratum was picked to be negatively related to the variation within the stratum. The correlations of the absolute value of the residuals with these reciprocals revealed that there was substantial heteroscedasticity present in the regressions. All regressions reported in the paper are unweighted and exhibited little evidence of heteroscedasticity.

The relationship of $(1-t_k)\alpha_k$ to net worth varies according to whether housing is included, and if included, how measured. For the narrowest definition of wealth exclusive of human wealth and homes, there is overall a significant positive relationship to net worth (see Table 4). A closer analysis, however, indicates that most of this positive relationship stems from households with net worth of less than $500,000 and that for households of greater means, the ratio shows a slight tendency to decrease (see Table 1). This behavior would be consistent with decreasing proportional risk aversion at least through net worth classes up to $500,000. This result is not affected appreciably if mixed assets were combined with risk-free rather than with risky assets. For the broader definition of wealth including housing as a risky asset but excluding human wealth, the ratio of $(1-t_k)\alpha_k$ is practically invariant to net worth if housing is measured by its equity value, and negatively related to net worth if housing is measured by its full value (see Table 4). These findings are consistent with constant proportional and increasing proportional risk aversion, respectively.[21] If housing is included as a risk-free asset, the tendency toward decreasing proportional risk aversion would be even greater than indicated by Table 1.

There are several potentially important biases in these findings. First, contrary to the assumptions of the model, individuals do not all hold the same portfolio of risky assets. Consider an individual whose portfolio of risky assets is perfectly correlated with the market portfolio of risky assets but is less volatile—that is, has a portfolio *beta* coefficient of less than 1.0. Since the risky portfolio of reference is the market portfolio, the calculated value of α_k would overstate the relative degree of risk in his portfolio. Likewise, if he were to hold a portfolio of risky assets with a *beta* greater than one, the calculated value of α_k would understate the degree of risk. The evidence in our 1975 paper indicates that households with lower incomes and presumably less wealth tend to hold portfolios of stocks with smaller portfolio *betas* than their richer counterparts although their portfolios are less well diversified.

Second, life insurance is not handled satisfactorily in this analysis since only the immediately realizable portion is treated as a risk-free asset. Life insurance does not conveniently fit into the dichotomy of risk-free and risky assets, and some other mode of analysis is required to handle this type of asset. However, the net surrender value of life insurance probably represents a significant understatement of the value of insurance as a device for reducing risk. Since life insurance is relatively much more important in the lower than in the upper net worth classes, correction for this deficiency of the model would again tend to change the results in the direction of decreasing proportional risk aversion.

Third, the calculated α_k might differ from its permanent value because of transitory elements. If a transitory reduction in net worth left unchanged the value of risky assets, the calculated α_k would be biased upwards. This situation might apply to an individual who was faced with a temporary reduction in net worth and because of transaction costs chose to reduce his risk-free assets. However, if he reduced only risky assets, the calculated α_k would be biased downwards. Since the former case is probably more prevalent for families of lower net worth and this type of transitory phenomenon is more likely to apply to these households, correcting for this possible bias would be expected to change the results in the same way as the previous bias.

---

[20] Though these comparative results indicate that the socioeconomic factors are roughly orthogonal to the logarithm of net worth, this does not mean that C_k^{-1} could not vary with these factors.

[21] The qualitative results are unchanged if households with closely held businesses are excluded. For instance, following the same order as in Table 4, the estimated coefficients on the logarithm of net worth would be: 0.076, −0.004, and −.250 with respective t-values of 8.4, −.61, and −11.48. These coefficients are roughly the same as those reported for the full sample. The regressions were also rerun with dummy variables included for age, education, and occupation. There is little change in the estimated coefficients on the logarithm of net worth.

Fourth, the model made the assumption that the tax rate on risky assets was the same as that on risk-free assets. In fact, the tax rate on risky assets tends to be less than that on risk-free assets. While the magnitude of this bias seems to be small, it results in an understatement of C_k.[22] However, it is not clear how this understatement varies with net worth.

To this point, the empirical analyses have ignored human wealth. The micro relationship (8) incorporates such wealth and can be used to examine the behavior C_k. Rearranging (8) yields

$$\text{(15)} \quad \frac{E(r_m - r_f)}{\sigma_m^2}\frac{1}{C_k} = (1 - h_k)(1 - t_k)\alpha_k + h_k(1 - t_k)\beta_{hk,m}$$

If $\beta_{hk,m}$ were known, the right-hand side of (15) could be used to analyze the relationship of C_k^{-1} to net worth. The variable $\beta_{hk,m}$, however, is not known, so that some other tack must be taken.

This other tack will be to assume the validity of two stochastic relationships and use these in conjunction with (15) to derive an empirically estimable regression. For the net worth class i, the first relationship takes the form

$$\text{(16)} \quad (C_k)^{-1} = (C^i)^{-1} + \eta_{ck}$$

where η_{ck} is a mean-zero disturbance independent of h_k and t_k, and C^i is the harmonic mean of C_k in the ith net worth class. This assumption is like the assumption of constant proportional risk aversion on the average, as embodied in (11), except that it only pertains to a single net worth class. Again for net worth class i, the second relationship takes the form

$$\text{(17)} \quad \beta_{hk,m} = \beta^i_{hm} + \eta_{hk}$$

where η_{hk} is a mean-zero disturbance independent of h_k and t_k, and β^i_{hm} is the mean of $\beta_{hk,m}$ in the ith net worth class.

The substitution of (16) and (17) into (15) with some simplification yields the regression

$$\text{(18)} \quad (1 - h_k)(1 - t_k)\alpha_k = \frac{E(r_m - r_f)}{\sigma_m^2}\frac{1}{C^i} - \beta^i_{hm}h_k(1 - t_k) + \mu_k$$

where μ_k is defined to be $[E(r_m - r_f)/\sigma_m^2]\eta_{ck} + h_k(1 - t_k)\eta_{hk}$. The independent variable, $h_k(1-t_k)$, and the dependent variable, $(1-h_k)(1-t_k)\alpha_k$, can be estimated from the cross-sectional data. The constant term allows an analysis of the relationship of the coefficient of proportional risk aversion to net worth class, and the slope coefficient gives an estimate of β^i_{hm}. It is easily verified that under the assumed stochastic processes, μ_k has an expected value of zero and is uncorrelated with the independent variable, so that the estimated coefficients will be unbiased.[23]

Using the total sample of households, regression (18) was first estimated for each of the six net worth classes inclusive of human wealth and with homes measured

---

[22] See fn. 37.

[23] The formula for μ_k would on the surface suggest that (19) suffers from some form of heteroscedasticity. Though this is true because of the way in which (17) and (18) were defined, it would be easy to redefine the variance of η_{hk} as a function of $h_k(1-t_k)$ so as to make (19) homoscedastic. Thus, the apparent heteroscedasticity of (19) is artificial. Further, the discussion of (19) at this point abstracts from measurement error in h_k.

TABLE 5—ESTIMATED COEFFICIENTS AND OTHER STATISTICS FOR REGRESSIONS OF THE FORM[a]

$$(1 - h_k)(1 - t_k)\alpha_k = \frac{E(r_m - r_f)}{\sigma_m^2}\frac{1}{C^i} - \beta^i_{hm}h_k(1 - t_k) + \eta_k$$

Net Worth Class[b]	Estimate of $\frac{E(r_m-r_f)}{\sigma_m^2}\frac{1}{C^i}$	Estimate of β^i_{hm}	$\bar{R}^2$	Unweighted Means $(1-t_k)\alpha_k$	Unweighted Means $(1-h_k)\cdot(1-t_k)\alpha_k$	Unweighted Means $h_k(1-t_k)$	Unweighted Means $(1-t_k)$	Number of Households
A. Total Sample[c]								
1–10	0.742	0.775	0.52	0.582	0.496	0.319	0.982	114
	(0.035)	(0.069)						
10–100	0.707	0.775	0.80	0.595	0.255	0.583	0.912	819
	(0.009)	(0.014)						
100–200	0.611	0.697	0.85	0.675	0.145	0.668	0.867	485
	(0.010)	(0.013)						
200–500	0.612	0.743	0.76	0.602	0.292	0.431	0.814	200
	(0.016)	(0.029)						
500–1000	0.529	0.613	0.44	0.584	0.412	0.192	0.705	90
	(0.020)	(0.072)						
over 1000	0.499	0.603	0.11	0.525	0.440	0.097	0.616	135
	(0.021)	(0.146)						
B. Households with Income from Salaries or Wages								
1–10	0.759	0.796	0.66	0.466	0.314	0.559	0.976	65
	(0.048)	(0.071)						
10–100	0.650	0.704	0.76	0.564	0.153	0.706	0.908	676
	(0.012)	(0.015)						
100–200	0.595	0.676	0.83	0.678	0.113	0.712	0.869	455
	(0.010)	(0.014)						
200–500	0.565	0.671	0.73	0.590	0.227	0.504	0.615	171
	(0.018)	(0.031)						
500–1000	0.567	0.712	0.60	0.615	0.387	0.254	0.714	68
	(0.022)	(0.070)						
over 1000	0.515	0.676	0.13	0.536	0.441	0.110	0.522	119
	(0.024)	(0.154)						

[a] See fn. for Table 1.
[b] Shown in thousands of dollars.
[c] Numbers in parentheses are standard errors.

by their equity value (see Table 5).[24] If unbiased, the constant terms indicate that the coefficient of proportional risk aversion on the average increases with net worth. The magnitude of the increase is not large. The estimate of $(C^i)^{-1}$ for those households with net worth in excess of $1 million is 33 percent less than the estimate for those households with net worth between $1 and $10 thousand. The estimates of β^i_{hm} tend to decrease as net worth increases starting at 0.775 for the lowest net worth class and ending at 0.603 for the highest net worth class.

There are several potentially important biases in these results. The first type pertain to violations of the assumptions used in developing (18). The second type pertain to limitations of the survey data. Turning to the first type, the assumptions implied that μ_k, which is a weighted sum of the deviations of $(C_k)^{-1}$ and $\beta_{hk,m}$ from their group means, is independent of $h_k(1-t_k)$ within a net worth class. This implication probably is not strictly true. For instance, households in the same net worth class with a lower proportion of their wealth in human capital might be expected to have lower tax rates which would tend to induce a correlation between η_k and $h_k(1-t_k)$. The possibility that the coefficient of proportional risk aversion might vary with socioeconomic factors would only cause a misspecification of (19) if these factors were correlated to h_k and t_k within a net worth class. Perhaps the most obvious socioeconomic factor which might be correlated with h_k and probably t_k within a class is the employment status of the members of the household. To determine the importance of this obvious misspecification, regression (18) was rerun on the sample of households reporting some income from salaries or wages. The results are not much different from the regressions for the total sample (see Table 5). For three out of six classes, the constant term was larger than before, while for the other three it was less, and the changes are not systematically related to net worth.

---

[24] If housing is measured instead by market value, the results are not much different. From lowest to highest net worth class, the constant terms would have been 0.810, 0.793, 0.634, 0.623, 0.530, and 0.500.

Let us now turn to the second type of biases which stem from limitations of the sample data. There are at least five principal problems. First, the balance sheets collected do not contain any estimate of the value of social security or pension fund benefits. Second, the value of relief payments is omitted. The importance of this omission is likely to be significant only in the lowest net worth class. Third, the value of human wealth may be misstated due to incorrectly assumed discount rates and growth rates.[25] Fourth, there may be transitory elements in the value of human capital. Fifth, the data do not allow an unambiguous determination of the return from closely held businesses attributable to wages. Taking the last problem first, (18) was reestimated for households with no closely held businesses. The results were unchanged.[26]

The first four problems all reduce logically to misstatements of the values of h_k and $\beta_{hk,m}$. In the model, social security and pension funds are assets over which a household has no control, so that formally these assets should enter into a more broadly construed h_k and would increase the value of this term. If it is assumed that the return on social security is uncorrelated with the return on the market portfolio of liquid risky assets, the inclusion of social security would reduce the value of a more broadly defined $\beta_{hk,m}$. Since pension funds are generally tied more directly to wage income, as a first approximation it is probably not too far from the mark to assume that their inclusion would not change $\beta_{hk,m}$ by much. The use of an incorrect discount rate or the inclusion of transitory elements would probably cause a misstatement in h_k but not necessarily in $\beta_{hk,m}$.

The total differential of (15) provides a way to measure the impact of these potential problems. Assuming dt_k and $d\alpha_k$ are zero in that t_k and α_k, derived from the household data, are known, the total differential takes the form

$$(19)\quad d\,\frac{E(r_m - r_f)}{\sigma_m^2}\,\frac{1}{C_k} = [(1 - t_k)\beta_{hk,m} - (1 - t_k)\alpha_k]dh_k + h_k(1 - t_k)d\beta_{hk,m}$$

The coefficients on $d\beta_{hk,m}$ can be estimated by net worth class as the means of $h_k(1-t_k)$ given in Table 5. Likewise from Table 5, the coefficients on dh_k by net worth class starting with the lowest can be estimated as: 0.179, 0.112, −0.071, 0.003, −0.152, and −0.154.

---

[25] The calculation of the value of human wealth made no adjustment for the probability of death before normal retirement. Formally, this is equivalent to assuming an incorrect discount rate and will be subsumed under this problem.

[26] From lowest to highest net worth class, the constant terms were estimated respectively as 0.711, 0.635, 0.541, 0.580, 0.522, and 0.449.

Replacing the total differentials in (19) by finite differences and recalling that h_k is measured as a proportion, it is seen that $|\Delta h_k|$ would have to be extremely large to have much effect upon the relationship of the market price of risk times $(C^i)^{-1}$ to net worth. Thus, the conclusions would appear robust to large errors in the assumed discount and growth rates.

Correcting for the omission of social security benefits may have an important impact and together with the biases previously discussed might easily change the conclusion to one of constant proportional risk aversion on the average. To determine the magnitude of such a correction, let us assume that the average value of social security benefits within a net worth class in 1962 was \$15,000.[27] Also assume that the value of social security does not change in any predictable way with changes in the value of the market portfolio, and that the net worth of the typical household in a net worth class is the midpoint except for \$2 million for the class with the largest net worth. On the basis of these assumptions, the estimates of $[E(r_m-r_f)/\sigma_m^2](C^i)^{-1}$ would change to 0.609 for the 1–10 thousand dollar class,[28] 0.583 for the 10–100 class, 0.561 for the 100–200 class, 0.592 for the 200–500 class, 0.521 for the 500–1,000 class, and 0.495 for the class of greatest wealth. These adjusted figures show considerably less evidence of increasing proportional risk aversion on the average. Assigning a larger value to social security benefits and adjusting for the other biases discussed above might even change the conclusion to decreasing proportional risk aversion.

To summarize, for what we regard as the most appropriate treatment of investment in homes and human wealth within the context of our model, the unadjusted results give some evidence of increasing proportional risk aversion on the average. Yet, the magnitude of this increase is not great. Corrections for the various deficiencies of the theoretical model, the econometric analyses, and the sample data would on balance tend to make the results closer to constant proportional risk aversion and might even produce evidence of moderately decreasing proportional risk aversion. Perhaps the most accurate single statement is: if there is any tendency for increasing or decreasing proportional risk aversion, the tendency is so slight that for many purposes the assumption of constant proportional risk aversion is not a bad first approximation.

IV. The Market Price of Risk

To assess numerical values for the coefficients of proportional risk aversion, the results in the previous section must be augmented with an assessment of the market price of risk, namely $E(r_m-r_f)/\sigma^2(r_m)$. This section will employ two different techniques to formulate such an assessment. The first will utilize realized rates of return on stocks and fixed interest obligations, while the second will attempt

---

[27] This estimate of value is based upon the following assumptions: The average household is headed by a person 45 years old and consists of a spouse and two children with average age of 10 years. Using the 1941 CSO table as in *CRC Standard Mathematical Tables* with a 2.5 percent interest rate and assuming maximum benefits with no increases in the future, the retirement portion was estimated roughly to be \$13,000 which represents an overstatement in view of the assumption of maximum benefits. The ancillary survivorship and disability benefits are more difficult to estimate but seem well under the 50 percent of retirement benefits as shown in government compilations. As a result, we have used an overall, probably somewhat conservative estimate of \$15,000 for the combined value of retirement and ancillary benefits. The use of 2.5 percent interest rate is consistent with the maintenance of the real value of the benefits over time.

[28] This number was calculated by noting that the inclusion of social security benefits would cause an increase of 0.487 in the value of h_k and a decrease of 0.690 in the value of β_{hm}. Applying these changes to (19) results in a decrease of 0.133, which when subtracted from 0.742 gives the adjusted estimate in the text.

TABLE 6—AVERAGE ANNUAL ARITHMETIC RATES OF RETURN AND STANDARD DEVIATIONS OF ANNUAL RETURNS 1872–1971

	New York Stock Exchange Stocks[a]				High Grade Corporate Bonds[b]				Calculated Market Price of Risk		
	Composite		Industrial						Stocks		
Period	Mean Annual Return	Standard Deviation	Mean Annual Return	Standard Deviation	Mean Annual Return	Standard Deviation	Bond[c] *Beta*	Risk-free[d] Rate	Composite	Industrial	Bonds
1872–1881	0.0953	0.177	0.0820	0.129							
1882–1891	0.0429	0.130	0.0720	0.117							
1892–1901	0.0879	0.118	0.0773	0.163							
1902–1911	0.0794	0.218	0.1111	0.318	0.0436	0.0391	0.136	0.0409	0.808	0.692	1.712
1912–1921	0.0532	0.195	0.1128	0.296	0.0415	0.0545	0.179	0.0495	0.097	0.721	−2.718
1922–1931	0.1020	0.268	0.1108	0.289	0.0572	0.0369	0.061	0.0447	0.798	0.789	9.226
1932–1941	0.0866	0.282	0.1255	0.348	0.0672	0.0376	0.081	0.0138	0.925	0.915	37.782
1942–1951	0.1725	0.136	0.1697	0.105	0.0244	0.0289	0.044	0.0127	8.673	14.136	14.084
1952–1961	0.1804	0.189	0.1859	0.203	0.0177	0.0456	−0.055	0.0315	14.165	3.740	−6.623
1962–1971	0.0742	0.130	0.0779	0.136	0.0259	0.0596	0.131	0.0512	1.372	1.447	−7.125
1902–1961	0.1124	0.216	0.1360	0.263	0.0419	0.0431	0.072	0.0322	1.716	1.504	5.249
1902–1971	0.1069	0.206	0.1277	0.249	0.0396	0.0457	0.078	0.0349	1.702	1.501	2.276
1926–1971	0.1155	0.212	0.1256	0.234	0.0381	0.0420		0.0232	1.963	1.788	8.487
1872–1971	0.0974	0.188	0.1125	0.221							

[a] From 1926, the figures are estimated from the Standard & Poor's Stock Indexes. For earlier years, the data are from *Common-Stock Indexes*. The Cowles *Index* was converted to the Standard & Poor's base by adjusting overlapping figures. The relative for the first quarter of the t-th year was taken to be $i_1=[P_1+\frac{1}{4}(D_1/P_1)P_1]/P_0$ where P_1 and D_1/P_1 are stock price and annualized dividend yield at end of quarter and the return for the year derived from the product of the four quarterly returns.

[b] Semiannual relatives were estimated as $(P+\frac{1}{2}C)/100$, where C is the coupon rate on a new 20-year bond sold at par and P is the price at the end of the 6 months. The coupon rate C was estimated as the yield to maturity given by Standard & Poor's for High Grade Corporate Bonds. The price P was calculated using the coupon rate C and the yield to maturity given by Standard & Poor's at the end of the 6 months on the assumption that such a yield would be appropriate for a bond with 19 years and 6 months to maturity. Multiplying successive pairs of these relatives yields estimates of the annual returns.

[c] The βs are derived from semiannual regressions of the form $r_b=\alpha+\beta r_m$ where r_b and r_m are annual rates of return on corporate high grade bonds and composite stocks, respectively.

[d] Prime Corporate Bonds 1-Year Maturity February Average from Sidney Homer, pp. 366–67, updated by Salomon Brothers.

to develop direct *ex ante* measures. Common stocks on the New York Stock Exchange (*NYSE*) will be used as the principal proxy for the market portfolio of risky assets partially because of the importance of these stocks in the actual market portfolio, but more pragmatically because of the lack of data on returns for most other risky assets. Nonetheless, the market price of risk will also be estimated using similar techniques with data from the bond market to confirm the estimates from the *NYSE* data.

For both industrial and all or composite *NYSE* common stocks, Table 6 shows for the past 100 years, 1872–1971, and selected subperiods, arithmetic averages of annual realized rates of return[29] as well as the standard deviations of these rates. The annual returns, which assume the quarterly reinvestment of dividends, are estimated from the work of the Cowles Commission prior to 1926 and, thereafter, from Standard and Poor's sources. The returns prior to 1926 may be less accurate. Over the past hundred years, the average annual return for the composite stocks was 9.7 percent and for the industrial stocks, 11.3 percent. The standard deviations of these annual returns were 0.19 and 0.22, respectively.

---

[29] These average rates of return would represent unbiased estimates of one-period holding returns if stationarity is assumed. To assess the sensitivity of these numbers to different horizons and since the arithmetic

Also included in Table 6 are corresponding mean annual returns and standard deviations for high grade corporate bonds. The only differences are that coupons are reinvested semiannually rather than quarterly and only the last 70 years are covered. The mean annual return on these bonds over the years 1902–71 was 4.0 percent which compares to a range of 10.7 to 12.8 percent for common stocks over the same period. In fact, in every one of the seven decades covered, stocks beat bonds and sometimes by substantial margins. However, the standard deviations of annual returns on bonds were always less than on stocks.

As a further check on the relationship of the bond and the stock markets, the semiannual returns for the bonds were regressed upon the corresponding returns for the *NYSE* composite stocks.[30] The slope coefficients or *beta* coefficients are displayed in Table 6 and indicate that the returns on bonds usually tend to fluctuate with those on stocks except in the 1952–61 decade when the coefficient is negative. Over the entire 70-year period, the *beta* coefficient is 0.078 which would imply, if the capital asset pricing model held, that the expected risk premium on bonds is 7.8 percent of that on stocks.[31]

Using the mean annual returns and the standard deviations calculated in the corresponding period, estimates of the market price of risk were derived. These estimates —as well as the risk-free rates—are contained in Table 6. The availability of a reliable series for the risk-free rate confined these estimates to the period after 1902. The estimates of the market price of risk over the longest period 1902 to 1971 were 1.7 for the composite stocks and 1.5 for the industrials, with not much different values for the 1902–61 period immediately preceding the date of the *FRB* survey. The estimate for the bond market was 2.3. It might be argued that these estimates are biased downwards because the variance of returns is estimated from a segment of the market of risky assets which would not reflect the full benefits from diversification. Additionally, assets which are both more and less risky than *NYSE* common stocks have been excluded in deriving the average risk premiums. It is difficult to assess the bias associated with these exclusions.[32]

The sampling errors in these estimates are undoubtedly large. To obtain some idea of the magnitude, the estimates of the market price of risk might be conceived of

---

mean raised to the appropriate *n*th power to obtain an *n*-year return can yield a seriously upward biased estimate of expected return over an *n*-year period (see Blume, 1974), an average of five-year returns over the 1872–1971 period (yielding a return relative of 1.587 for the composite index) was compared with the fifth power of the average annual returns (1.592), with little difference in results.

[30] For a discussion of the specification of this type of regression, the reader is referred to Blume (1970).

[31] Over the 1902–71 period, the average annual risk premium on *NYSE* composite stocks was 0.0720. Multiplying this figure by 0.078 would yield an expected risk premium for bonds of 0.0056, which compares to the actual average annual risk premium on bonds of 0.0047.

[32] The market price of risk was also estimated using, instead of variance of returns, the variance of the difference of the annual return and the risk-free rate. If the risk-free rate were constant over time, the two variances would, of course, be the same. These alternative estimates were marginally lower. For the period 1902–71, this alternative estimate was, for instance, 0.04 smaller for the composite stocks than that presented in Table 6.

as drawings from some underlying probability distribution. Using this approach, the average of the seven decade estimates for the composite stocks is 2.4 which is somewhat larger than the estimate derived from the total period reflecting the large sampling error and possibly some nonstationarities in the distribution over time. The standard deviation of this average is 1.2. If the market price of risk were normally distributed, this would mean that there is an 0.88 probability that the true value of this parameter is greater than 1.0. The corresponding average and standard deviation for the industrials are 3.2 and 1.9, respectively.

An alternative way to assess the market price of risk is to develop *ex ante* measures of the expected returns on risky assets and the variance of those returns. To this end, *ex ante* measures of expected return were developed by adding the current yield to an estimate of the future growth rate of earnings per share.[33] The future growth rate was estimated by the arithmetic and by the geometric mean of the annual growth rates over the prior 10 years.[34] For the 70 years ending in 1971, the mean estimate of the expected annual return for the composite stocks ranged from 13.0 to 9.0 percent, depending upon the way in which the future growth rate of earnings was estimated.[35] It might be recalled that the corresponding mean of annual realized returns was 10.7 percent. For comparison, the yield to maturity on high grade corporate bonds is also given in Table 7.

These estimates of expected rates of return for composite stocks were used to derive an estimate of the market price of risk for each of the 70 years from 1902 through 1971. Subtracting the risk-free rate appropriate to the year from these expected returns yielded 70 risk premiums, which were then divided by estimates of the variance of the return for these stocks. These variances were estimated from the 10 realized annual returns immediately preceding the year for which the expected return was assessed. Table 7 presents the average of these 70 estimates of the market price of risk as well as averages for selected subperiods. The average covering the 70 years ending in 1971 was 3.1 using the arithmetic estimate of *ex ante* expected return and 2.0 using the geometric, which are again greater than 1.0 though somewhat higher than the figures derived from realized returns.

To obtain some feeling for the magnitude of the sampling errors associated with these averages for the overall period, the standard deviation of the mean was calculated using every tenth estimate of the market price of risk from 1901 through 1971. Using every estimate instead of every tenth would have resulted in a downward biased estimate of the standard deviation unless explicit account was taken of the fact that successive estimates are based in part upon the same data. The standard deviation of the mean was 0.99 for the market price of risk based upon the arithmetic rate of growth in earnings and 0.87 for the geometric rate of growth. Assuming these standard deviations of the mean can be applied to the means of the 70 annual estimates (though there is reason to believe the true standard deviations would be less), the probability that the market price of risk using arithmetic rates of growth is greater than 1.0 would be 98

---

[33] Implicit in this formulation are the assumptions: 1) Earnings grow at a constant rate in perpetuity. 2) The payout ratio and discount rates are constant.

[34] Estimates using only 5 prior years of data gave unreasonably low rates of return in the early 1920's and middle 1930's which with the appropriate dividend yield would have implied expected returns as low as −22 percent. As long as cash is available, no investor if he really held this belief would hold stock. Using 10 years of prior data yielded negative estimates of the expected market return in only a few cases and then only with the geometric.

[35] This range is consistent with that found in a survey of stockholders by W. G. Lewellen, R. C. Lease, and G. C. Schlarbaum.

TABLE 7—"EXPECTED" ANNUAL RATES OF RETURN FOR COMMON STOCKS AND BONDS
1902–71

Period	High Grade Corporate Bonds Yields to Maturity	New York Stock Exchange Composite Stocks		Calculated Market Price of Risk for Stocks Using	
		Arithmetic Estimate[a]	Geometric Estimate[b]	Arithmetic Estimate	Geometric Estimate
1902–1911	0.0452	0.157	0.132	6.090	4.674
1912–1921	0.0519	0.132	0.097	1.989	1.120
1922–1931	0.0486	0.177	0.059	3.639	0.319
1932–1941	0.0363	0.064	0.011	0.518	−0.042
1942–1951	0.0270	0.160	0.126	3.829	3.075
1952–1961	0.0362	0.137	0.122	4.472	3.843
1962–1971	0.0562	0.086	0.080	1.228	1.047
1902–1961		0.138	0.091	4.423	2.165
1902–1971	0.0431	0.130	0.090	3.109	2.005

[a] The Arithmetic Estimate is the arithmetic mean of annual expected returns where the expected return for the t-th year is taken to be $i_t = D_t/P_t + GE_t$, where GE_t = arithmetic mean of previous 10 years of earnings growth. Dividends yields were obtained and earnings growth rates calculated from Standard and Poor's Quarterly "Earnings, Dividends and Price Earnings Ratios."

[b] The Geometric Estimate is the same except that GE_t is the geometric mean of the previous 10 years of earnings growth.

percent. The corresponding probability for the market price of risk based upon geometric rates of growth would be 87 percent.

This section has used several different approaches to estimating the market price of risk. All of these methods indicate that the market price of risk is greater than 1.0. The probabilities attached to this conclusion were assessed and, while not always at the frequently cited 5 percent level of significance, were consistent with a high level of significance. To the extent that there is some independence among the various estimates, the probability that the market price of risk is greater than 1.0 would be considerably larger than those reported for each of the individual estimates.

V. The Synthesis

The empirical results in Section III indicate that the assumption of constant proportional risk aversion for households is a fairly accurate description of the market place. More specifically, the study found that with a plausible treatment of investment in homes the net worth of a household had little value in explaining the level of its coefficient of relative risk aversion. Before taking account of potential biases, there was a slight tendency of this coefficient to increase with net worth. After adjusting for the biases, little such tendency remained.[36]

The main exception to this finding of constant proportional risk aversion occurred in the analysis based upon an extremely narrow definition of wealth which excluded homes, the associated mortgages, and human wealth. In this case, there was evidence of decreasing proportional risk aversion. This tendency would be more pronounced if homes were considered a risk-free asset. Since an investor probably considers the value of his home and at least to some extent his human wealth in formulating his investment decisions, the analyses based upon broader definitions of wealth are more pertinent in assessing how the coefficient of proportional risk aversion varies with net worth.

---

[36] These conclusions do not preclude the possibility that the coefficient of proportional risk aversion varies with socioeconomic factors which are uncorrelated with net worth.

In addition, Section III derived various estimates of the product of the market price of risk and the reciprocal of the coefficient of proportional risk aversion. Regardless of how broadly wealth was defined, the estimate of this product was, with one unimportant exception, always less than 1.0. Though the value obtained depends on the definition of wealth, the product of the market price of risk times the reciprocal of the coefficient of proportional risk aversion appears to be between 0.5 and 0.8.

Coupling this conclusion with an estimate of the market price of risk allows an assessment of the coefficient of relative risk aversion. The evidence in Section IV showed with a high level of significance that the market price of risk is in excess of 1.0. The implication is that the coefficient of relative risk aversion for the typical household is in excess of 1.0—contrary to the properties of the *log* utility function. Since the market price of risk is probably around 2.0 or more, the coefficient of proportional risk aversion is more likely to be in excess of two.[37]

Households, of course, only represent part of the demand for risky marketable assets. The aggregate demand function would have to take into account the impact of other investors such as eleemosynary institutions and financial intermediaries. Regardless of how their utility or, more accurately, decision functions are derived, the fact that the ratios of risky assets to net worth for these institutions are usually less than 1.0 and not uncom-

---

[37] This conclusion is based upon the assumption that the tax rates on all sources of income are the same when in fact tax rates differ. The effect is that the estimates of the coefficients of proportional risk aversion reported in the text are downward biased, although the bias is not large and does not vary much among net worth classes. With differential taxes and not including human wealth, an estimate of $\alpha_k C_k$ would be

$$\frac{(1-t_m)E(r_m)-(1-t_f)r_f}{(1-t_m)^2\sigma_m^2}$$

where t_m is the tax rate on risky liquid assets and t_f the tax rate on risk-free assets. Assuming the same tax rate on both kinds of assets, the estimate of $\alpha_k C_k$ would be given by

$$\frac{E(r_m)-r_f}{(1-t_c)\sigma_m^2}$$

where t_c is a weighted average of t_f and t_m. Noting that $(1-t_m)>(1-t_f)$ and $E(r_m)>r_f$, the following inequality holds

$$\frac{(1-t_c)(1-t_m)}{(1-t_m)^2}E(r_m)-\frac{(1-t_c)(1-t_f)}{(1-t_m)^2}r_f>E(r_m)-r_f$$

Upon dividing by $(1-t_c)\sigma_m^2$, the inequality shows that the estimate of $\alpha_k C_k$ derived under the assumption of the same tax for both kinds of assets is biased downwards.

If human wealth is included and assumed taxed at the same rate t_{hw} as risk-free assets, $\alpha_k C_k$ would be given by the same formula as the first one in the footnote times the quantity

$$\left[(1-h_k)+\frac{h_k}{\alpha_k}\cdot\frac{1-t_f}{1-t_m}\cdot\beta_{hk\cdot m}\right]^{-1}$$

If the same rate is applied to all assets, $\alpha_k C_k$ is given by the second formula times the quantity

$$\left[(1-h_k)+\frac{h_k}{\alpha_k}\cdot\beta_{hk\cdot m}\right]^{1}$$

Since $(1-t_f)<(1-t_m)$, not taking account of differential taxes will result in a downward biased estimate of the second term in the product. Since the first term is subject to the same biases as before, the net effect will be that estimates of C_k which do not adjust for the differential taxes associated with human and risk-free wealth will be downward biased.

Under reasonable estimates of the relevant variables, it can be shown that the assumption that tax rates on all sources of income are the same biases downward the coefficient of proportional risk aversion for the broadest definition of wealth by about 8 percent for the \$10,000–\$100,000 net worth class and by 9 percent for the \$500,000–\$1,000,000 class. These results assume: $E(r_m)=0.10$, $r_f=.05$, and $\sigma_m^2=.025$ for all households; for the low net worth households, $t_f=t_{hw}=.10$, $t_m=.02$ in view of the heavy weight of housing in the marketable risky assets of this group, and $t_c=.08$; and for the high net worth households, $t_f=t_{hw}=.35$, $t_m=.25$ in view of the heavy weight of unincorporated business and the small weight of housing and tax-exempts.

monly considerably less than 1.0 would seem to imply that their coefficients of relative risk aversion would tend to be at least 2.0. If, as is probably correct, the riskiness of institutional portfolios as a whole has remained less than that of households, their coefficients of relative risk aversion would tend to be greater than those of households.

In concluding, it is worthwhile to review some of the more important limitations of the models under which the empirical data were interpreted. Such limitations point the way to further work. First, the model made no adjustment for inflation. The presence of unanticipated inflation would mean that no asset denominated in nominal dollars could be considered risk free. Second, the model, in assuming that the changes in the value of an asset were continuous, could not properly incorporate insurance. Third, the model did not incorporate adequately the unique characteristics of housing. Fourth, the model assumed that an investor made his portfolio choice in a one-period context. Thus, the model, in contrast to a multiperiod model, is not rich enough to capture the possibilities of hedging against future changes in the investment opportunity set.[38] For households, perhaps the most widely available hedges are homes which protect against future changes in the cost of housing.

[38] Merton has developed a theoretical multiperiod model which does incorporate continuous changes in the investment opportunity set. E. F. Fama and J. D. MacBeth have developed some preliminary tests of the economic content of Merton's model. It might be noted that a utility function characterized by constant proportional risk aversion is consistent with myopic behavior so long as investment yields in different periods are serially independent.

REFERENCES

K. J. Arrow, *Essays in the Theory of Risk-Bearing*, Chicago 1971.

M. E. Blume, "Portfolio Theory: A Step Towards Its Practical Application," *J. Bus., Univ. Chicago*, Apr. 1970, *43*, 152–73.

———, "Unbiased Estimators of Long-Run Expected Rates of Return," *J. Amer. Statist. Assn.*, Sept. 1974, *69*, 634–38.

——— and I. Friend, "The Asset Structure of Individual Portfolios and Some Implications for Utility Functions," *J. Finance*, May 1975, *30*, 585–603.

A. P. Budd and R. H. Litzenberger, "The Market Price of Risk, Size of Market and Investor's Risk Aversion: A Comment," *Rev. Econ. Statist.*, May 1972, *54*, 204–06.

J. Crockett and I. Friend, "Consumer Investment Behavior," in R. Ferber, ed., *Determinants of Investment Behavior*, Nat. Bur. Econ. Res. Conference Report, New York 1967, 15–127.

E. F. Fama and J. D. MacBeth, "Tests of the Multiperiod Two-Parameter Model," *J. Finance Econ.*, May 1974, *1*, 43–66.

I. Friend, "Rates of Return on Bonds and Stocks, the Market Price of Risk and the Cost of Capital," in Friend and J. Bicksler, eds., *Risk and Return in Finance*, Cambridge 1975.

M. J. Gordon, G. E. Paradis, and C. H. Rorke, "Experimental Evidence on Alternative Portfolio Decision Rules," *Amer. Econ. Rev.*, Mar. 1972, *62*, 107–18.

N. H. Hakansson, "Capital Growth and the Mean-Variance Approach to Portfolio Selection," *J. Finance. Quant. Anal.*, Jan. 1971, *6*, 517–57.

S. Homer, *The History of Interest Rates*, New Brunswick 1963.

Y. Landskroner, "The Determinants of the Market Price of Risk," unpublished doctoral dissertation, Univ. Pennsylvania 1975.

H. A. Latané, "Criteria for Choice Among Risky Ventures," *J. Polit. Econ.*, Apr. 1959, *67*, 144–55.

W. G. Lewellen, R. C. Lease, and G. C. Schlarbaum, "Patterns of Investment Strategy and Behavior Among Individual Investors," mimeo., Purdue Univ. 1974.

J. Lintner, "The Market Price of Risk, Size of Market and Investor's Risk Aversion," *Rev. Econ. Statist.*, Feb. 1970, *52*, 96–99.

H. Markowitz, "Investment for the Long Run," in I. Friend and J. Bicksler, eds., *Risk and Return in Finance*, Cambridge 1975.

D. Mayers, "Non-marketable Assets and Capital Market Equilibrium Under Uncertainty," in M. C. Jensen, ed., *Studies in the Theory of Capital Markets*, New York 1972.

A. H. Meltzer, "The Demand for Money: The Evidence from the Time-Series," *J. Polit. Econ.*, June 1963, *71*, 219–46.

R. Merton, "An Intertemporal Capital Asset Pricing Model," *Econometrica*, Sept. 1973, *41*, 867–87.

J. Mossin, "Security Pricing and Investment Criteria in Competitive Markets," *Amer. Econ. Rev.*, Dec. 1969, *59*, 749–57.

R. Roll, "Evidence on the Growth-Optimum Model," *J. Finance*, June 1973, *28*, 551–66.

S. A. Ross, "Uncertainty and the Heterogeneous Capital Goods Model," *Rev. Econ. Stud.*, forthcoming.

P. A. Samuelson, "The Fundamental Approximation Theorem of Portfolio Analysis in Terms of Means, Variances, and Higher Moments," *Rev. Econ. Stud.*, Oct. 1970, *37*, 537–43.

——— and R. Merton, "Fallacy of the Log-Normal Approximation to Optimal Portfolio Decision-Making Over Many Periods," in I. Friend and J. Bicksler, eds., *Risk and Return in Finance*, Cambridge 1975.

S. C. Tsiang, "The Rationale of the Mean-Standard Deviation Analysis, Skewness Preference, and the Demand for Money," *Amer. Econ. Rev.*, June 1972, *62*, 354–71.

Cowles Commission for Research in Economics, *Common-Stock Indexes*, 2d ed., Bloomington 1939.

CRC Standard Mathematical Tables, 11th ed., Cleveland 1957.

Internal Revenue Service, *Statistics of Income*, Washington 1962.

An Intergenerational Model of Population Growth

By ASSAF RAZIN AND URI BEN-ZION*

The problem of population growth has been only partially formulated in an intertemporal context. Therefore, a link between population theory and economic growth theory is still missing.[1] This paper attempts to provide such a link by analyzing an intergenerational model of optimum population growth. The model accommodates the idea that decisions on increases in population should consider the new population's quality of life. This concept is elaborated upon in a static framework by Gary Becker and Gregg Lewis, Dennis De Tray and Robert Willis. Other literature on population growth, such as Harold Votey, disregards this consideration altogether.[2] More explicitly, this paper assumes that the utility of each generation is a function of the level of its consumption and the *number* and *utility* of the newly born people. We are thus led to consider the utility of an infinite number of generations. This model differs from Paul Samuelson's intergenerational model of pure consumption in which the utility of each generation depends only on that generation's own consumption.

Recently, Philip Neher (1971) analyzed a model of fertility in primitive economies; it is based on the assumption that parents use children as investments to insure future consumption needs. Neher's analysis focuses entirely on the pension motive for having children;[3] our model assumes that fertility is determined so as to balance the parents' welfare derived from additional children against that derived from the children's quality of life.

We first outline a general model of optimum population growth. In order to discuss explicitly various economic aspects of population, we use a simple case in the analysis. Section I sets up the basic model. Section II analyzes the short-run equilibrium conditions. Section III discusses the long-run equilibrium conditions. Section IV presents an explicit solution for a simple case which yields a constant rate of population growth. Section V explores the effects on population growth of public support for investment in children. Section VI analyzes the relationship between population growth and the value of time. Section VII considers intergenerational transfers and pensions. Section VIII relates the analysis to real world population and policy problems.

I. The Model

Let there be identical individuals; let t be a point in time, and also the name of the generation of people born at t−1. Let L_t be the number of people of generation t. Let c_t be their per capita lifetime consumption, and let U_t be the utility indicator of

---

* Departments of economics, Tel-Aviv University and Technion, Israel, respectively. This paper was prepared while we were visiting at the department of economics, University of Minnesota. We are indebted to Marc Nerlove, T. Paul Schultz, Theodore W. Schultz, to the editor, and to the referees. However, we alone are responsible for errors remaining in the paper. Partial financial support from A.I.D. through the Center of Economic Development at the University of Minnesota is acknowledged.

[1] For recent surveys of economic approaches to population, see Marc Nerlove and T. Paul Schultz.

[2] See the authors for a critical discussion of the formulation in Votey.

[3] On this subject, see Warren Robinson's criticism and Neher's reply.

their preferences. Each generation lives one period.[4] Thus, L_{t+1} (the number of people of generation t+1) may also be regarded as the number of people born at t. Let $\lambda_t = L_{t+1}/L_t$ be the per capita number of people born at t; let K_t be the total amount of physical capital, and let $k_t = K_t/L_t$ be the per capita amount at t. In general, when U_t is a function of c_t, λ_t, and U_{t+1}, we can write an intergenerational utility indicator $V = U_t(c_t, \lambda_t, U_{t+1}(c_{t+1}, \lambda_{t+1}, U_{t+2}(\ldots)))$, where $t = 0, 1, 2, \ldots$.[5]

A few remarks on the welfare function $V(\ )$ are in order. First, note that in effect the first generation lives forever through its offspring; therefore, our approach is more general than the one adopted by Neher (1971). According to Neher's approach, parents' concern for the future is represented by the utility of consumption enjoyed by their children but only during the children's first period of life. In our model, the concern of parents for their children is not limited at all.

Second, note that since in this paper we use $V(\ )$ as the guide for intergenerational allocation of society's resources, it is only the first generation's welfare that directly enters into the social welfare function.[6]

We assume that preferences are the *same* for each generation and can be represented by an *additive* utility function $U_t(c_t, \lambda_t) + \beta U_{t+1}$; this implies

$$(1) \qquad V = \sum_{t=0}^{\infty} \beta^t U(c_t, \lambda_t)$$

where β is the subjective factor by which the current generation discounts the utility of the next generation.[7] Note the following consistency of the model: under full certainty, the parent generation's plan for future generations' increases in consumption and population is identical with that which the future generations will desire for the given amount of capital they will inherit. This is obviously the result of the specific form of utility function in (1). Let the capital of the next generation be produced according to a linear homogeneous function $F(\cdot)$,[8]

$$(2) \qquad K_{t+1} = F(K_t - C_t, L_t)$$

where $K_t - C_t$ is the total saving. In per capita terms, (2) can be rewritten as:

$$(3) \qquad \lambda_t k_{t+1} = f(k_t - c_t)$$

where $f(k_t - c_t) \equiv F(k_t - c_t, 1)$.

Note that the budget constraint in equation (3) implies that for generation t, the amount of k_t (which is inherited from previous generations and is *given*) can be allocated to two ends: to consumption c_t, or to the amount of resources to be left over to each individual in the next generation k_{t+1}. These quantities vary according to changes in population λ_t. This is analogous to the budget constraint used in Becker and Lewis where λ is regarded as the number of children in a family and k_{t+1} as the amount invested in their "quality."[9]

---

[4] This assumption implies nonoverlapping of generations and therefore excludes basic mechanisms such as loans rather than intergenerational gifts. In Section VII we show that our results carry over to a more general model in which this assumption is relaxed.

[5] The inclusion of population growth in the social utility function also has an empirical implication for the measurement of welfare improvement. Growth of per capita income by itself is an inappropriate measure of welfare improvement, and it is biased against countries with a high rate of population growth.

[6] A more general approach is to define a social welfare function for intergenerational equity which includes as its argument a vector of welfare functions belonging to different generations. This is a complex problem in both its conceptual and analytical aspects, and lies beyond the scope of the paper. Our approach should be regarded as a first attempt in this area.

[7] In particular, note that the rate of discount is constant. Allowing in the model for a variable rate of discount will not change any qualitative results of the paper.

[8] The reader may refer to Nissan Liviatan for an analysis of optimum economic growth which utilizes the discrete-time formulation.

[9] In the model, the inherited capital may include both human and nonhuman capital (when they are regarded as perfectly substitutable in production).

The decision problem of the current generation can now be written as:

$$V(k_0) = \max \sum_{t=0}^{\infty} \beta^t U(c_t, \lambda_t) \tag{4}$$

$$0 \leq c_t \leq k_t; \qquad 0 \leq \lambda_t \leq \bar{\lambda}$$

subject to (3) and a given level of k_0 where $\bar{\lambda}$ is the maximum feasible level of population growth. Marginal utilities are positive and diminishing.[10]

We can now provide the following characterization of an interior solution of (4).

$$\frac{\partial U}{\partial \lambda}(c_t, \lambda_t) = \frac{\beta}{\lambda_t} k_{t+1} \frac{\partial U}{\partial c}(c_{t+1}, \lambda_{t+1}) \tag{5}$$

$$\frac{\partial U}{\partial c}(c_t, \lambda_t) = \frac{\beta}{\lambda_t} \frac{\partial f}{\partial k}(k_t - c_t) \frac{\partial U}{\partial c}(c_{t+1}, \lambda_{t+1}) \tag{6}$$

Equation (5) may be interpreted as describing the optimum decision with respect to the level of population growth λ_t. On one hand, an extra unit of λ_t will increase welfare by the marginal utility of population growth, the left-hand side of (5). On the other hand, from (3), for given values of k_t and c_t, this increase in λ_t will reduce the next generation's per capita capital k_{t+1} by the amount of k_{t+1}/λ_t; this entails a loss for the next generation of (k_{t+1}/λ_t) times the marginal utility of consumption. This utility loss (discounted by β) appears on the right-hand side of (5). In equilibrium, according to equation (5), the gain and the loss in utility are equated.

Equation (6) may be interpreted as describing the optimizing level of consumption c_t. On one hand, an extra unit of c_t will increase utility by the marginal utility of consumption, the left-hand side of (6). On the other hand, this entails a reduction in the next generation's capital; this loss is expressed in utility terms by the right-hand side of (6).

II. The Short Run

In the short run, when the initial value of k is given, the system of equations (3), (5), and (6) may be solved for the quantities of the three "goods": current consumption c_0, population growth λ_0, and next-generation capital k_1 (all expressed in per capita terms). The term k_1 does not appear directly in the utility function and thus is not an ordinary good; we must use its imputed value. Since at equilibrium an extra unit of capital can be used either for consumption or for saving with equal marginal gain (equation (6)), we have[11]

$$\frac{\partial V(k_t)}{\partial k} = \frac{\partial U(c_t, \lambda_t)}{\partial c} \tag{7}$$

Equation (7), specialized for t = 1, will yield the imputed value of the next-generation capital k_1.

Figures 1 and 2 represent a short-run interior solution, where marginal rates of substitution and marginal rates of transformation between the three goods are equalized. At point B in Figure 1, the marginal rate of substitution between λ and k_1

$$\frac{\partial U(c, \lambda)}{\partial \lambda} \bigg/ \beta \frac{\partial V(k_1)}{\partial k}$$

is equal to k_1/λ, the marginal rate of transformation between λ and k_1. This condition can be derived by substitution of (7) into (5). As shown in Figure 1, a necessary condition for an interior solution for λ and k_1 is that the slope of the indifference curve diminishes faster than the slope of the

---

[10] In the real world one may conceive of situations where the marginal utility of a child is nonpositive. For example, in a cooperative system such as the Israeli kibbutz where parents do not directly bear the cost of raising their children, parents may reach a point of saturation. We can reinterpret $\bar{\lambda}$ as the level of λ where $\partial U/\partial \lambda = 0$.

[11] It can be shown that $V(k)$ is concave, i.e., $\partial^2 V/\partial k^2 > 0$.

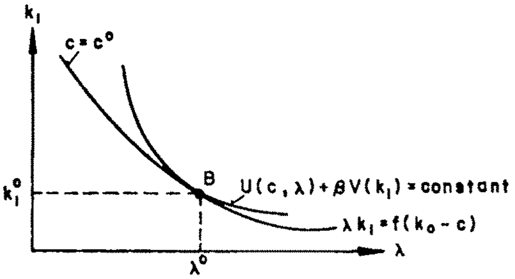

k_1
$c = c^0$
k_1^0
B
$U(c, \lambda) + \beta V(k_1) = \text{constant}$
$\lambda k_1 = f(k_0 - c)$
λ^0
λ

FIGURE 1

transformation curve as λ increases. This implies the following condition:[12]

$$(8)\quad \left(\frac{\lambda U_{\lambda\lambda}(c,\lambda)}{U_\lambda(c,\lambda)}+1\right)+\left(\frac{k_1 V_{kk}(k_1)}{V_k(k_1)}+1\right)<0$$

where U_λ, V_k, $U_{\lambda\lambda}$, and V_{kk} are first- and second-order partial derivatives.

At point C in Figure 2, the marginal rates of substitution and transformation between c and λ—given respectively by $[\partial U(c,\lambda)/\partial c]/[\partial U(c,\lambda)/\partial\lambda]$ and $f'(k_t - c)/k_1$—are equalized. This condition can be derived from equations (5) and (6). In general, not all three "goods" (c, λ, and k_1) need be normal. In the diagrams we observe different cross-sections of a three-dimensional diagram; therefore when k_0 is increased, there are shifts in both indifference and transformation curves in Figures 1 and 2, so that the expansion paths need not be upward sloping and the increase in k_0 may result in a decline in the quantity of one or two of these goods.

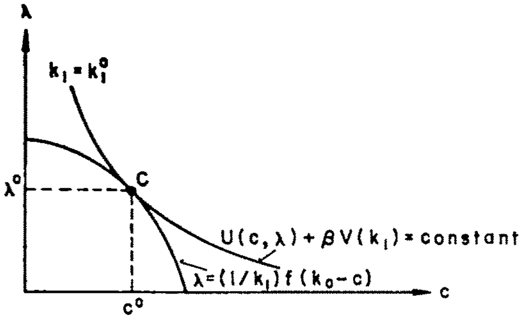

λ
$k_1 = k_1^0$
λ^0
C
$U(c, \lambda) + \beta V(k_1) = \text{constant}$
$\lambda = (1/k_1) f(k_0 - c)$
c^0
c

FIGURE 2

Putting this differently, in the budget constraint (3), either λ or k_1 may be regarded as a price of the other. In the event that k_0 increases and λ and k_1 change, the wealth effect is mixed with the price effect. As a result, the direction of change in the quantities of all subsets of the three goods is not known a priori. This point was elaborated upon by Becker and Lewis in the context of their model.

III. The Long Run

At steady states (when they exist), the system of equations (3), (5), and (6) reduces to:

$$(9)\qquad \lambda k - f(k - c) = 0$$

$$(10)\qquad \lambda U_\lambda(c,\lambda) - \beta k U_c(c,\lambda) = 0$$

$$(11)\qquad \lambda/\beta - f'(k - c) = 0$$

where $U_\lambda = \partial U/\partial\lambda$, $U_c = \partial U/\partial c$.

Equation (11) may be interpreted as the "modified Golden Rule," where the marginal productivity of capital $f' - 1$ is equal to the rate of population growth n plus the subjective rate of discount δ.[13]

One remarkable property of the steady state is that it is invariant to shifts in utility which keep constant the marginal rate of substitution between population growth and consumption U_λ/U_c. Therefore, the steady-state equilibrium (but not the path converging to it) results from an *ordinal* utility (with a given discount factor β) since any monotone increasing transformation of $U(\ )$ leaves the steady state unchanged.

---

[12] Note that one of the cases in which an interior equilibrium is not achieved is when the utility function is logarithmic, i.e., $U(c, \lambda) = a\ log\ c + b\ log\ \lambda$. In this case, the marginal rate of substitution between k_1 and λ is $a\lambda/\beta b k_1$, while the marginal rate of transformation is λ/k_1. For example, equality is achieved if $a = \beta b$, but then the indifference curve coincides with the transformation curve, leaving the values k_1 and λ indeterminate. Otherwise, i.e., when $(a \neq b\beta)$, a solution may not exist. To see this, consider a two-period model with $V = a\ log\ c_0 + b\ log\ \lambda_0 + \beta\ log\ k_1$. We can rewrite this as $V = a\ log\ c_0 + \beta\ log\ (\lambda_0 k_1) + (b - \beta)\ log\ \lambda_0$. The product $\lambda_0 k_1$ as such appears explicitly in the budget constraint $\lambda_0 k_1 = f(k_0 - c_0)$; thus, when $b - \beta < 0$, λ_0 may be regarded as a free good with a negative marginal utility. However, $\lambda_0 = 0$ is not a solution. Therefore, a solution does not exist.

[13] Note that $\lambda = 1 + n$, $\beta = 1/(1+\delta)$. Therefore, $f' - 1 = (\lambda/\beta) - 1 = (1+n)(1+\delta) - 1 \approx \delta + n$.

IV. The Constant Rate of Population Growth

Since it is difficult to further characterize the general solution, it may be instructive to consider a simple case. We believe that the analysis of this case enables one to understand the essential characteristics of the model, and that it can be used to make inferences in more general situations.

Let the utility function be given by:

$$U(c, \lambda) = a \log c + v(\lambda) \tag{12}$$

where a is a positive number and $v(\lambda)$ is the utility arising from population increase with positive and diminishing marginal utility. In order to guarantee an interior solution, we assume that the elasticity of the marginal utility of population increase is greater than one (i.e., $-\lambda v_{\lambda\lambda}/v_{\lambda} > 1$).

Let the production function be of a Cobb-Douglas form

$$f(k - c) = r(k - c)^{\alpha} \tag{13}$$

where the constants r and α are positive and $\alpha \leq 1$. The intertemporal utility function which is obtained upon substituting (12) into (1) has two special properties. First, the marginal rate of substitution between consumptions at different points of time is not affected by population increases. Second, the elasticity of substitution between consumptions at different points of time is unitary. It turns out that these are the factors which determine in the following analysis the particular results. However, it seems to us that the functions of consumption and population growth which are derived from (12) and (13) are not necessarily inconsistent with those observed in the real world.

Substituting (12) and (13) into (6), we find a solution:

$$c_t = (1 - \alpha\beta)k_t \tag{14}$$

Substitution of (14) into (3) and (5) yields

$$\lambda_t k_{t+1} = r(\alpha\beta)^{\alpha} k_t^{\alpha} \tag{15}$$

$$\lambda_t v_{\lambda}(\lambda_t) = \frac{a\beta}{1 - \alpha\beta} \longrightarrow \lambda = \lambda^* \tag{16}$$

We therefore obtain a consumption function with a unitary-income elasticity (equation (14)) and a constant rate of population growth λ^* (equation (16)). The next-generation capital stock will increase as current resources increase (equation (15)).

Solving the difference equation (15) for $\log k_t$ using (14) and (16), we get

$$\log k_t = \log\left[r(\alpha\beta)^{\alpha}/\lambda^*\right] \cdot \left[\frac{1 - \alpha^{t+1}}{1 - \alpha}\right] + \alpha^t \log k_0 \tag{17}$$

Equations (14), (16), and (17) yield a complete short-run solution for our problem. The comparative statics properties of the short-run solution are straightforward. For example, an increase in the discount rate of the utility of future generations will increase the present consumption and the number of children which are undiscounted in the utility of the present generation; it will also reduce the capital per child which affects the utility function only indirectly through the effect on the utility of the next generation. If individuals are risk averters an increase in the intergenerational discount factor may result from increases in uncertainty regarding the continuity of the family, namely, uncertainty associated with the length of the horizon of the extended family.[14] Also, it is seen from (16) that an increase in the capital productivity parameter α will increase intergenerational accumulation of capital and lower the rate of population growth.

---

[14] For a detailed analysis of uncertainty regarding the horizon of the decision-making unit, see Menachem Yaari.

Turning to the long run, for values of α below unity we have a stable equilibrium k^* (as $t \to \infty$),

$$k^* = [r(\alpha\beta)^\alpha/\lambda^*]\, 1/(1-\alpha) \tag{18}$$

V. Public Support for Investment in Children and Population Growth

One characteristic of the modern state is the existence of subsidies for the education of the young generation, paid by taxes imposed on the parent generation. From the individual parent's point of view, there is no necessary relationship between the subsidies (which are received on a per child basis) and tax payments (which are paid on the basis of the parents' income). The model of Section IV is now viewed as describing the behavior of a typical (extended) family. We analyze the effect of government policy on population growth by using this model.

Subsidies for the investment in children born at t can be represented by a proportional increase in the amount of capital invested in a child, k_{t+1}. Equation (3) can therefore be rewritten as

$$\lambda_t k_{t+1} = (1+s) f(k_t - c_t) \tag{19}$$

where s is the rate of the subsidy.

Assume that the public subsidy is financed by a proportional tax τ, so that the disposable wealth at point t is $k_t(1-\tau)$. In order for the government's budget to be balanced over the infinite horizon and for the government spending not to exceed its revenue at any point in time, we need

$$\sum_{t=0}^{T} \lambda_t k_{t+1} s \leq \sum_{t=0}^{T} \tau k_t, \qquad \text{for } T = 1, 2, \ldots \tag{20}$$

We assume that the tax-cum-subsidy policy, when enacted, is expected by individuals to last forever. What is the effect of once-and-for-all changes in the rate of subsidy, with a balanced government budget, on population growth? In the presence of the tax-cum-subsidy policy, equations (5) and (6) are modified as follows.

$$U_\lambda(c_t, \lambda_t) = \frac{\beta}{\lambda_t}(1-\tau) k_{t+1} U_c(c_{t+1}, \lambda_{t+1}) \tag{5'}$$

$$U_c(c_t, \lambda_t) = \frac{\beta}{\lambda_t} f'((1-\tau)k_t - c_t) \cdot (1+s) U_c(c_{t+1}, \lambda_{t+1}) \tag{6'}$$

In the special case considered in Section IV, these equations can be solved explicitly to get

$$c_t = (1-\tau-\alpha\beta) k_t \tag{21}$$

$$\lambda_t v_t(\lambda_t) = \frac{(1-\tau)\alpha\beta}{1-\tau-\alpha\beta} \to \lambda_t = \lambda^*(\tau) \tag{22}$$

$$log\, k_t = log\,[r(1+s)(\alpha\beta)^\alpha/\lambda^*(\tau)] \cdot \frac{1-\alpha^{t+1}}{1-\alpha} + \alpha^t log\,(1-\tau)k_0 \tag{23}$$

where s and τ are chosen so as to satisfy (20).

From equation (21) it can be inferred that consumption is negatively related to the rate of the income tax τ, whereas it is unaffected by the subsidy. Since the elasticity of the marginal utility of population increase is assumed to exceed one,[15] and the term on the right-hand side of (22) is positively related to τ, an increase in the tax rate τ lowers the rate of population growth λ, whereas the rate of population growth is not affected by the subsidy. It is the Hicks-neutral nature of the subsidy that is the reason for the insensitivity of both λ and c to changes in the subsidy rate. Putting this somewhat differently, absolute prices of λ and c in all periods change by the same proportion leaving relative prices unchanged, thus creating no incentives for changes in these variables.

---

[15] See the discussion in Section IV.

The model thus implies that an increase in the subsidy to the next generation's capital which is financed by an increase in the rate of income tax will lead, as expected, to lower levels of consumption and population growth and to an increase in the rate of capital accumulation.

VI. Population Growth and the Value of Time

The model of the preceding sections did not capture one important aspect of the population problem, namely, the effect of changes in the value of time on the optimal policy with regard to population increase. In this section, time requirements for raising children are explicitly introduced. Population increases will imply reductions in the amount of time devoted to productive purposes.

Let g denote the *fraction* of the lifetime of the parents' generation devoted to raising children, which is assumed to be an increasing function of the number of children per family.[16] Inverting this relationship, the number of children per family λ_t is an increasing function $\phi(\ )$ of the fraction of time which the parents' generation devotes to raising children g_t.

$$\lambda_t = \phi(g_t), \quad \phi'(\ ) > 0 \tag{24}$$

The total time devoted by the parents' generation to productive activities is $(1-g_t)L_t$. This new element will require the following modifications in the equations of Section I:

$$\lambda_t k_{t+1} = (1 - g_t) f\left(\frac{k_t - c_t}{1 - g_t}\right) \tag{25}$$

$$U_\lambda(c_t, \lambda_t) = \beta \frac{k_{t+1}}{\lambda_t} U_c(c_{t+1}, \lambda_{t+1}) + \frac{\beta F_L}{\phi'(g_t)\lambda_t} U_c(c_{t+1}, \lambda_{t+1}) \tag{26}$$

where $F_L = f - (k-c)/(1-g)f'$ is the marginal productivity of labor.

Equation (26) differs from equation (5) by having an additional term on its right-hand side. This term corresponds to the cost of time associated with the population increase. One extra unit of λ requires an increase of the magnitude $1/\phi'(g)$ in g (see (24)). The resulting decrease in labor force participation will lower the per capita production of capital of the next generation according to the marginal productivity of labor divided by the rate of population growth F_L/λ. The utility value of the loss viewed by the present generation is $\beta F_L U_c/\lambda\phi'$. It is instructive to solve this model explicitly in the special case considered in Section IV. The modification of the solution in Section IV is in equation (16) which determines the rate of population growth λ while equation (14) is unaffected. We get

$$\phi(g_t) v_\lambda(\phi(g_t)) = \frac{a\beta}{1 - \alpha\beta} + a\beta \frac{1 - \alpha}{1 - \alpha\beta} \frac{\phi(g_t)}{\phi'(g_t)} \longrightarrow g_t = g^* \tag{27}$$

Equation (27) implies a *constant* fraction of time g^* to be devoted to raising children. This implies, in view of (24), a *constant* rate of population growth.

The reader may recall that the model assumed constant returns to time invested in raising children. A straightforward implication of the analysis is that if the function relating the number of children to total time devoted to raising children exhibits increasing (diminishing) returns to scale, the rate of population growth will increase (decrease) with the capital-labor ratio.

This model suggests that in relation to population growth, it is misleading to measure the value of time by the marginal productivity of labor as such. The marginal productivity of labor varies over time, together with the capital stock (except at a steady state). The factor which does *not* vary with the capital stock in this special case (and hence does not vary over time) is the *utility value of time:* $U_c F_L$. This is so since a proportional increase in the capital stock will lead to an equiproportionate increase in the consumption of the next generation, and, as a result, to an equiproportionate decrease in its marginal utility. Also, the marginal productivity of labor goes up by the same proportion. This leaves both the value of time $U_c F_L$ and the rate of population growth λ unchanged. In general, however, the value of time which determines, among other things, the rate of population growth might move either with or in opposition to the capital stock.[17] In particular, development of the economy over time (associated with the accumulation of capital) need not imply an increasing time trend for the utility value of time.[18]

---

[16] This can be interpreted as maternity costs. Investment in the human capital of children can be introduced in a similar way.

VII. Intergenerational Transfers and Pensions

In previous Sections I–III the model assumed no overlapping of generations, in the sense that parents were not alive during their children's productive time. In this section we relax this assumption in order to accommodate basic mechanisms such as intergenerational loans, which parents can utilize to delay consumption from productive to nonproductive periods; this is in addition to intergenerational transfers which were considered previously. While this approach is in line with that of Neher (1971), there are two major differences. First, in the model considered by Neher, children are the sole means of income transfer and pension. In our model, there exists a durable good by which parents' saving can be transferred from one period to another. Second, Neher imposes the restriction that total product is equally shared by productive and nonproductive people. In our model, the distribution of the total product between generations is endogenous.

Consider the family unit at a point in time t composed of three generations; the productive parents generation and the children, denoted by superscript 1; and the old generation who have retired and perform no economic functions, denoted by superscript 2. Generalizing the simple model of the previous sections, at time t parents choose the number of children λ_t, current consumption for its members and children c_t^1, the parents' next period consumption c_{t+1}^2, and the amount of capital transfer k_{t+1} (expressed in per capita units). The problem is to maximize

$$\sum_0^\infty \beta^t U(c_t^1, c_{t+1}^2, \lambda_t) \tag{28}$$

subject to $\lambda_t k_{t+1} = f(k_t - c_t^1 - c_t^2/\lambda_{t-1})$

The first-order conditions are

$$\frac{\partial U}{\partial c^1}(c_t^1, c_{t+1}^2, \lambda_t) = f'(k_t - c_t^1 - c_t^2/\lambda_{t-1}) \frac{\partial U}{\partial c^2}(c_t^1, c_{t+1}^2, \lambda_t) \tag{29}$$

$$\frac{\partial U}{\partial c^1}(c_t^1, c_{t+1}^2, \lambda_t) = (\beta/\lambda_t) f'(k_t - c_t^1 - c_t^2/\lambda_{t-1}) \cdot \frac{\partial U}{\partial c^1}(c_{t+1}^1, c_{t+2}^2, \lambda_{t+1}) \tag{30}$$

---

[17] Correspondingly, its partial effect on population growth is negative or positive.

[18] This is analogous to the effect of changes in the wage rate on labor-leisure choice for an individual consumer. Changes in the wage rate have both substitution and income effects. Similarly, in our model, changes in the capital-labor ratio will have substitution and income effects on population and consumption. While the substitution effect is negative the full effect of capital-labor ratio on population growth is not known a priori. The negative effect on fertility caused by the change in the value of time of mothers which is discussed in Schultz, should be regarded mainly as a substitution effect.

$$\lambda_t \frac{\partial U}{\partial \lambda}(\overset{1}{c}_t, \overset{2}{c}_{t+1}, \lambda_t) = \frac{\partial U}{\partial c^2}(\overset{1}{c}_t, \overset{2}{c}_{t+1}, \lambda_t) \cdot [f(k_t - \overset{1}{c}_t - \overset{2}{c}_t/\lambda_{t-1}) - \overset{2}{c}_{t+1}] \tag{31}$$

Consider again a simple case

$$U(\overset{1}{c}_t, \overset{2}{c}_{t+1}, \lambda_t) = a \log \overset{1}{c}_t + b \log \overset{2}{c}_{t+1} + v(\lambda_t) \tag{32}$$

$$f(k_t - \overset{1}{c}_t - \overset{2}{c}_t/\lambda_{t-1}) = r(k_t - \overset{1}{c}_t - \overset{2}{c}_t/\lambda_{t-1})^{\alpha}$$

where b/a can be regarded as the intrageneration discount factor. Substituting (32) into (29)–(31) we can solve for the desired level of consumption and population. This is given in (33).

$$\begin{aligned} \overset{1}{c}_t &= m(k_t - \overset{2}{c}_t/\lambda_{t-1}) \\ \overset{2}{c}_{t+1} &= p k_{t+1} \lambda_t \\ \lambda_t v'(\lambda_t) &= b \frac{1-p}{p} \end{aligned} \tag{33}$$

where $m = 1 - \alpha(a\beta + b)/\beta(a + \alpha b)$

$p = b(1 - \alpha\beta)/(a\beta + b)$

The consumption functions in (33) have unitary income elasticities and population growth λ_t is *constant*. From (33), pension per family of the old (nonproductive) generation t, c^2_{t+1}, can be written as follows.

$$\overset{2}{c}_{t+1} = \lambda_t pr(k_t - \overset{2}{c}_t/\lambda_{t-1})^{\alpha}(1 - m)^{\alpha} \tag{34}$$

It can be seen from (34) that the pension is an increasing function of the rate of population growth λ_t, and the amount of saving made in the previous period by the now nonproductive generation, $(1-m) \cdot (k_t - c^2_t/\lambda_{t-1})$. Thus, the pension of the nonproductive generation is determined by the number of productive families and by the amount of transfers made by such a family. The major difference between Neher's results and ours pertains to the latter.

In Neher's model the portion of saving of a productive family that is received by nonproductive families depends on a given distributional rule. In this model it is affected by savings in the previous period of the nonproductive generation. In the model we do not specify the mechanism which determines the share of output received by the nonproductive generation. In particular, this income need not take the form of an intergenerational gift; in fact, it can consist of interest payments on intergenerational loans and returns to ownership of physical capital.[19] For simplicity the model did not consider the possibility of children caring for their parents. Obviously, this factor by itself will increase pensions.

VIII. Conclusion and Implications

The model has a few implications for real world problems which we now wish to emphasize. First, the increase in capital productivity tends to increase intergenerational accumulation of capital and to lower the rate of population growth. This may be a result of capital-biased technological change.[20]

Second, we discuss a government program of subsidizing investment in children's future well-being, financed by income tax. Specifically, the type of subsidy considered is such that government shares with parents the investment in children. Obviously, parents can use a subsidy either to increase the investment per child with a given number of children, or to increase the number of children with a given amount invested in a child, or a combination of the two. The analysis suggests that the subsidy by itself will tend to increase investment in a child, and has a weaker effect on the number of children. However, the effect of the system of subsidies financed by taxes is to reduce the rate of population growth as well as the level of consumption, thereby increasing capital (per capita) of the next generations.

---

[19] It may also be noted that if there are preferences as to the form of the pension income, it may affect the allocation of investment between human and nonhuman capital.

[20] The transfer of technology from developed to underdeveloped countries through technical assistance may lead to technical change of this kind.

Third, we show that as long as time requirements of raising a child are given, the increase in capital per worker (which in turn raises real wages) has no clear effect on the rate of population growth since it involves both income and substitution effects. The same holds true for Hicks-neutral technical change. However, the partial effect of labor-biased technical change which raises the utility value of time is to lower the rate of population growth. Obviously, if technical change is associated with a reduction in the time requirement of raising children, it will have a positive effect on population growth. Also, we show that in the case of economies of scale in raising children, the rate of population growth is positively related to the level of capital per worker; it is negatively related to the level of capital per worker in the case of diminishing returns.[21]

Finally, the case of overlapping generations is incorporated into the model as an extension of the recent contribution by Neher (1971). The extension allows the pension of the nonproductive generation to depend on its savings, in addition to being dependent on the number of its productive offspring. The model implies that the pension of the now nonproductive generation will be positively related to the amount of capital per worker in their productive period, to the productivity of capital, and to the intragenerational and intergenerational discount factors.

The model has also been extended in two additional directions: differences in tastes among generations and uncertain population changes.[22] The main results of these extensions (not presented here) are the following:

Considering the effects of differences in tastes between generations on current decisions regarding consumption and population changes, the analysis suggests that a consumption-oriented trend over generations tends to lower the current rate of population growth in addition to its negative effect on current consumption. This is similar to the effect of a change in the factor by which the utility of future generations is discounted on these variables. Changes in the utility function pertaining directly to children have a weaker effect on current population growth and consumption.

It is shown that increased uncertainty with regard to the continuity of the extended family (which is equivalent to increased intergenerational discount rates) tends to increase the rate of population growth, whereas it lowers the rate of capital accumulation over generations. Another kind of uncertainty in the form of stochastic disturbances to survival rates of new members of the family has a detrimental effect on population growth.

---

[21] Cooperative systems which have economies of scale in raising children, such as the Israeli kibbutz, have relatively greater numbers of children per family for a given level of capital per capita. It is suggested by the model that an increase in capital per worker (over time) will raise the rate of population growth in these systems.

[22] A draft containing the derivations of these results is available on request from the authors.

REFERENCES

G. S. Becker and H. G. Lewis, "On the Interaction between the Quantity and Quality of Children," *J. Polit. Econ.*, Mar./Apr. 1973, suppl., *81*, S279–88.

U. Ben-Zion and A. Razin, "The Number and Quality of Children," mimeo., 1972.

D. N. De Tray, "Child Quality and the Demand for Children," *J. Polit. Econ.*, Mar./Apr. 1973, suppl., *81*, S70–95.

N. Liviatan, "A Diagrammatic Exposition of Optimal Growth," *Amer. Econ. Rev.*, June 1970, *60*, 302–08.

P. A. Neher, "Peasants, Procreation, and Pensions," *Amer. Econ. Rev.*, June 1971, *61*, 380–89.

———, "Peasants, Procreation, and Pensions: Reply," *Amer. Econ. Rev.*, Dec. 1972, *62*, 979.

M. Nerlove, "Household and Economy: Toward a New Theory of Population and Economic Growth," *J. Polit. Econ.*, Mar./Apr. 1974, suppl., *82*, S200–18.

W. C. Robinson, "Peasants, Procreation, and Pensions: Comment," *Amer. Econ. Rev.*, Dec. 1972, *62*, 977–78.

P. A. Samuelson, "An Exact Consumption-Loan Model of Interest With or Without the Social Contrivance of Money," *J. Polit. Econ.*, Dec. 1958, *66*, 467–82.

T. P. Schultz, "A Preliminary Survey of Economic Analyses of Fertility," *Amer. Econ. Rev. Proc.*, May 1973, *63*, 71–78.

T. W. Schultz, "The High Value of Human Time: Population Equilibrium," *J. Polit. Econ.*, Mar./Apr. 1974, suppl., *82*, S2–10.

H. L. Votey, "The Optimum Population and Growth: A New Look," *J. Econ. Theory*, Sept. 1969, *1*, 273–90.

R. J. Willis, "A New Approach to the Economic Theory of Fertility Behavior," *J. Polit. Econ.*, Mar./Apr. 1973, suppl., *81*, S14–64.

M. E. Yaari, "Uncertain Lifetime, Life Insurance and the Theory of the Consumer," *Rev. Econ. Stud.*, Sept. 1965, *32*, 137–50.

Search in the Labor Market and the Duration of Unemployment: Some Empirical Evidence

By John M. Barron*

Recently several writers have examined job search behavior by an unemployed worker.[1] In the analysis, the period of time between successive wage offers to an individual has commonly been taken as constant. John McCall, for instance, assumed that an individual invariably obtained one job offer per period. Dale Mortensen (1970b) equivalently assumed a constant probability of receiving a wage offer each period. In this context, job search has been characterized as search for an acceptable wage.[2] The implication is that the expected length of unemployment depends solely on the individual's acceptance (or reservation) wage, optimally chosen, and the existing wage offer distribution.

One aspect of job search heretofore not emphasized is the individual locating a vacancy (wage offer). This element of search is captured below by introducing a random visiting of firms by the job searcher to locate an employment opening.[3] This extension permits presentation of a more complete job search theory which, as Charles Holt (1975) suggests, makes allowance for the time it takes to search firms for vacancies as well as the time it takes to search vacancies for suitable wages. The approach allows explicit consideration of an "involuntary" aspect of search not present in the existing literature.

Section I develops a model of job search by an individual that incorporates the above aspects of search. This model then allows in Section II a fruitful discussion of the elements affecting an individual's acceptance wage and the duration of unemployment. It becomes clear that the expected duration of unemployment entails more than a comparison of an acceptance wage and the wage offer distribution. Indeed, the model makes possible computation of a proxy for the probability that the typical individual engaged in job search will accept a wage offer. It is found that

---

* Assistant professor, Purdue University. This research was partially supported by a Baker-Weeks Fellowship provided through the Brookings Institution and a doctoral dissertation grant (No. 91-44-74-39) from the U.S. Department of Labor, Manpower Administration. I am indebted to Herschel Grossman and John Kennan for helpful comments on earlier drafts of this paper. In addition, I benefitted from comments by an anonymous referee. However, the usual disclaimers hold.

[1] For example, see articles by McCall, Mortensen (1970a,b, 1974), Reuben Gronau, Michael Rothschild (1974), and Robert Lucas, Jr. and Edward Prescott. Kenneth Burdett offers an excellent review of several results of such efforts. In addition, Rothschild (1973), focusing on market equilibrium, has provided a survey of "... recent theoretical work characterizing markets whose participants act on the basis of sketchy and incomplete information" (p. 1283).

[2] The analysis considers expected income maximizing behavior by the unemployed job searcher. Some recent models of job search have replaced expected income maximizing behavior with expected utility maximizing behavior in a nontrivial manner (see John Seater or John Danforth). This alternative approach suggests inclusion of such factors as the nonpecuniary aspects of job offers, leisure, consumption, and risk in the individual's optimal search policy.

[3] This job search method of going directly to an employer is the predominant method of job search according to recent surveys (see *Employment and Earnings*).

this variable varies directly with the duration of unemployment. Section III discusses the importance of this finding, particularly as it relates to Phillips curve theory. The conclusions of this paper provide support for the arguments of Donald Parsons, and Robert Barro and Herschel Grossman that question the relevance of emphasizing "employment-acceptance friction" in job search to explain the observed relationship between unemployment and aggregate demand.

I. A Simple Search Model

Consider an economy consisting of n_t firms in which each firm requires various types of labor inputs. For such an economy, examination of a division of the labor market by occupation is possible. Though there would exist within each firm a potential distribution of job opportunities, each firm is assumed to have no more than one vacancy for any one of the q types of labor at any time t. For the economy, the number of vacancies for each of the different types of labor required as inputs is designated by $v_t^1, \ldots, v_t^i, \ldots, v_t^q$. The vacancies are assumed randomly distributed among the n_t firms each period. The total number of vacancies in the economy is denoted by v_t.

Existing in the economy is a stock of unemployed workers, denoted by u_t, who are engaged in job search. Each potential employee is considered to possess a distinct combination of previous experience, skills, and formal education that identifies the particular type of labor for which he (she) is suited. Each searcher may then be said to be correlated with one "occupation." The number unemployed in the ith occupation is denoted by u_t^i.

The job searcher is operating in an environment of uncertainty both with respect to the location of a job for which he is qualified and the specific wage, given a distribution of wages, attached to a particular job. In this setting, one can conceptually distinguish search to find a vacancy from search induced by the inequality between an offer and an acceptance wage.

In one period, an unemployed worker expects to contact k firms. A period is defined such that k is sufficiently small ($k<1$) so that the probability of more than one searcher in the ith occupation contacting the same firm in a given period approaches zero.[4] For the individual, the average time (number of periods) taken to visit one firm is $1/k$. The probability, in a random search procedure, of locating a vacancy on contacting a firm is v_t^i/n_t for the ith occupation in period t. Then, in any period, the probability of finding a vacancy is $v_t^i k/n_t$.[5] To simplify notation, let

$$\theta_t^i = v_t^i k/n_t \tag{1}$$

The expected number of periods before an individual in the ith occupation finds a vacancy is $1/\theta_t^i$.

Let C^i denote the monetary (direct) cost of search per unit of time spent searching for the representative individual in the ith occupation. This cost would include transportation expenses and other expenditures.

Once having found a vacancy, the individual must decide whether to accept the wage offer w_t^i, or continue to search. The optimal rule is characterized by an acceptance (or reservation) wage (see Morris DeGroot 1968, 1970). For the representative individual in the *i*th occupation in period t, let $f^e(w_t^i)$ denote the perceived common density function of each of the observations w^i.[6] Then the expected return from searching for one more period R_t^i is given by

---

[4] A similar assumption, in a different context, is made by Gronau. Alternatively, the same result would follow from the assumption that u_t^i is sufficiently small relative to n_t.

[5] Holt suggests that search "proceeds from more promising areas of search to successively less promising ones" (1970, p. 228). This idea has been used by S. C. Salop in part to explain declining acceptance wages. *Ex ante* information about firms and the corresponding systematic job search could be characterized in our model by a falling rather than constant θ_t^i for an individual. The present analysis abstracts from such an effect.

$$R_t^i = \theta_t^i H^i \int_{a_t^i}^{\infty} (w_t^i - a_t^i) f^e(w_t^i) dw_t^i \tag{2}$$

where a_t^i is the acceptance wage of the representative individual and H^i is the expected length of employment in the next job. The searcher stops when he receives a wage offer at least as large as a_t^i, where a_t^i is the unique solution of the equation:[7]

$$C^i = R_t^i \tag{3}$$

That is, the acceptance wage is chosen to equate the marginal cost of searching for one more period C^i, with the expected marginal return from continued search R_t^i. In turn, R_t^i consists of the expected gain on locating another wage offer times the probability of locating a wage offer in the next period θ_t^i.

Note that the above analysis includes no discounting. Consider the case of an infinite time horizon but a positive time discount rate s. Then the cost of continued search includes the value of time as an important element in addition to monetary outlays.[8] In particular, equation (3) becomes

$$C^i + a_t^i = \theta_t^i (1/s) \int_{a_t^i}^{\infty} (w_t^i - a_t^i) f^e(w_t^i) dw_t^i \tag{4}$$

The right side of equation (4) represents the discounted expected return from searching for one more period. The left side of (4) is the cost of continued search which now includes foregone income. As before, the acceptance wage a_t^i is chosen to satisfy equation (4). The cost of continued search may be further altered to the extent that unemployment compensation is available or to the extent that search activity is considered by the individual to be different from work activity.

From equation (3) or (4) it can be shown that an increase in H^i, the expected length of employment, or a decrease in the discount rate will increase the acceptance wage of the representative individual by increasing the expected return to search. Similarly, a recognized increase in v_t^i, the number of vacancies in the *i*th occupation, or k (a measure of search efficiency) will increase the acceptance wage by raising the return to search. A decrease in C^i will increase the acceptance wage of the representative individual in the *i*th occupation a_t^i by decreasing the cost of continued search. A rightward shift in $f^e(w_t^i)$ will also increase a_t^i.

II. The Duration of Unemployment: Empirical Evidence

For the *i*th occupation, the mean probability that a wage offer is accepted, P_t^i, is

$$P_t^i = \int_{a_t^i}^{\infty} f(w_t^i) dw_t^i \tag{5}$$

where $f(w_t^i)$ is the actual wage offer distribution. For the entire economy, the average probability P_t that a searcher accepts a job which he has found is

$$P_t = \sum_{i=1}^{q} (u_t^i / u_t) P_t^i \tag{6}$$

The main interest of this paper is in aggregate phenomena. To simplify aggregation, assume the q occupations to be identical in the variables of interest. Specifically, assume that $P_t^i = P_t$ and that $\theta_t^i = \theta_t = v_t k/n_t q$, where θ_t is the probability of locating a vacancy in period t for the individual. Equivalently, one may consider the *i*th occupation as the representative occupation. As a consequence, the aggregate analysis to follow is the same as an analysis of the *i*th occupation and distributional effects are avoided.

---

[6] In order to avoid unnecessary complications, assume that the individual considers himself knowledgeable of the probability distributions associated with job search. This assumption precludes adaptive behavior arising from search when the distributions are unknown. Rothschild (1974) has examined one case in which the wage offer distribution is unknown.

[7] It is required that the variance of $f^e(w_t^i)$ be finite.

[8] A finite time horizon similarly changes the cost of search per unit time from that formerly proposed. In addition, a finite time horizon implies a declining acceptance wage for an individual engaged in job search (see Gronau).

The importance of determining P_t is that in part it determines the average duration of unemployment D_t, and thereby the unemployment rate U_t of the economy. Robert Flanagan, among others, has shown that in a steady state,

$$U_t = F_t D_t \tag{7}$$

where F_t is the rate of flow of individuals into unemployed status. The model presented in this paper allows examination of the factors involved in determining D_t. Specifically, given θ_t, the probability of locating a vacancy in period t, the average duration of unemployment in the economy is

$$D_t = 1/\theta_t P_t \tag{8}$$

Economic fluctuations are commonly characterized by movements in the number of vacancies. The question of how such a change in v_t affects equation (8) may now be addressed. Assume a fall in the number of vacancies. Consider the effects of such a change on three different parameters involved in determining the average duration of unemployment.

EFFECT 1: According to equation (1), a decrease in v_t increases the expected length of search time to find a vacancy, $1/\theta_t$. Given the average probability of accepting a wage offer P_t, this effect means, by equation (8), a rise in the average duration of unemployment.

EFFECT 2: According to equation (2), a perceived decrease in v_t by decreasing θ_t decreases the expected return from additional search for individuals in each occupation.[9] By equation (3) or (4), this effect means a reduction in acceptance wages. Given a distribution of wage offers, this reduction in acceptance wages implies an increase in P_t (equation (5)), providing to some degree an offsetting force for decreasing the average duration of unemployment.

EFFECT 3: A fall in v_t may be accompanied by a leftward alteration in the distribution of wage offers $f(w_t)$. According to equation (5), to the extent that individuals do not respond by reducing their expectations concerning wage offers $f^e(w_t^i)$, and thereby their acceptance wages, this effect reduces P_t and thus increases the expected duration of unemployment. It is of interest to note that to the extent that individuals fail to recognize a reduction in the number of vacancies as well as in the wage offer distribution, then it is more likely that Effect 3 will dominate Effect 2 regarding P_t.

An explicit expression of the net change in the average duration of unemployment D_t, due to a change in the number of vacancies using either equation (3) or (4) is provided in the Appendix. The net change refers only to the first two effects mentioned above. However, the sign of this expression is unclear. Fortunately, it is possible using existing data to obtain a measure of the effect changes in v_t have on both P_t and D_t. The method for computing D_t and P_t follows.

The probability that, in a given period, the representative unemployed individual in the economy finds and accepts employment is $\theta_t P_t$.[10] The distribution of the number of periods before employment occurs is geometric of the form.

---

[9] Since $1/\theta_t$ is the expected number of periods it takes to locate a vacancy, a reduction in v_t may equivalently be interpreted (by transposing θ_t to the left side of equation (3) or (4)) as increasing the expected cost of search for another wage offer.

[10] The analysis abstracts from variations in the rate at which unemployed workers leave the labor force.

$$(9)\quad g(\mathrm{T};\theta_t P_t) = (\theta_t P_t)(1-\theta_t P_t)^{\mathrm{T}-1}\qquad \mathrm{T}=1,2,3,\ldots$$

The probability that employment occurs within the first T periods, $Pr(E \leq \mathrm{T})$, is then

$$(10)\quad Pr(E \leq \mathrm{T}) = \sum_{h=1}^{\mathrm{T}} (\theta_t P_t)(1-\theta_t P_t)^{h-1} = 1-(1-\theta_t P_t)^{\mathrm{T}}$$

There exist data recording the number of people unemployed for different lengths of time. In general, denote G_t^{ab} and $G_{t-\mathrm{T}}^{ab}$ as the number of people who have been unemployed for at least a periods but for not more than b periods at time t and time t−T, respectively. The variables G_t^{ab} and $G_{t-\mathrm{T}}^{ab}$ would have no common elements but would be adjacent in the sense that $b-a=\mathrm{T}$. Let $G_{t+\mathrm{T}}^{bc}$ denote the number unemployed for at least b periods but for not more than c periods at time t+T. If $c-b$ equals $b-a$, then the ratio $G_{t+\mathrm{T}}^{bc}/G_t^{ab}$ corresponds to $1-Pr(E\leq \mathrm{T})$ which by equation (10) equals $(1-\theta_t P_t)^{\mathrm{T}}$. If $c-b$ equals two times $b-a$, then a quadratic equation results of the form

$$(11)\quad G_{t+\mathrm{T}}^{bc} = G_t^{ab}(1-\theta_t P_t)^{\mathrm{T}} + G_{t-\mathrm{T}}^{ab}(1-\theta_t P_t)^{2\mathrm{T}}$$

Solving and rearranging,

$$(12)\quad \theta_t P_t = 1 - \left[\frac{-G_t^{ab}+\left[(G_t^{ab})^2+4G_{t+\mathrm{T}}^{bc}\,G_{t-\mathrm{T}}^{ab}\right]^{1/2}}{2G_{t-\mathrm{T}}^{ab}}\right]^{1/\mathrm{T}} = 1-j(G_{t-\mathrm{T}}^{ab},\,G_t^{ab},\,G_{t+\mathrm{T}}^{bc})$$

The particular case relating various unemployed groups when $c-b=2(b-a)$ is presented in detail above since estimates of the variables G_t^{ab}, $G_{t-\mathrm{T}}^{ab}$, and $G_{t+\mathrm{T}}^{bc}$ in equation (12) can be obtained from data reported in the *Employment and Earnings* publication of the Bureau of Labor Statistics (*BLS*). Specifically, $G_{t-\mathrm{T}}^{ab}$ and G_t^{ab} were approximated by the seasonally adjusted number unemployed less than five weeks in the survey week of months one (t−T) and two (t), respectively. The variable $G_{t+\mathrm{T}}^{bc}$ then was given by the seasonally adjusted number unemployed for at least five weeks but for not more than fourteen weeks in the survey week of month three (t+T); T was set equal to four. Substituting these values into equation (12), an estimate of $\theta_t P_t$ is obtained. The actual estimate of $\theta_t P_t$ is an average about time t. Since $D_t = 1/\theta_t P_t$, equation (12) thus provides the method for computing an estimate of D_t. Figure 2 graphs D_t so computed for the period April 1969 to November 1973.[11]

Figure 1 presents the "mean duration of unemployment" statistic published by the *BLS*, $\underline{D}_t$, to indicate the fact that the computed D_t is less than $\underline{D}_t$, as has been reported elsewhere (see Hyman Kaitz). While the computed D_t measures the expected duration of unemployment for an individual becoming unemployed in period t, $\underline{D}_t$ measures the average time currently unemployed individuals have been searching for employment. Steve Salant has shown that the difference between $\underline{D}_t$, the mean of the "interrupted spell length," and D_t may be explained under the assumption of varying probabilities of accepting a wage offer among individuals. Further, the exact relation between $\underline{D}_t$ and D_t will vary as the distribution of P_t varies among individuals. Another factor influencing $\underline{D}_t$ but not D_t is changing flows of people entering the unemployed state.

The method used to compute P_t follows from equation (12). Given that $\theta_t = v_t k/n_t q$, then

---

[11] The monthly series for $G_{t-\mathrm{T}}^{ab}$, G_t^{ab}, and $G_{t+\mathrm{T}}^{bc}$ used were for the periods of March 1969 to October 1973, April 1969 to November 1973, and May 1969 to December 1973, respectively.

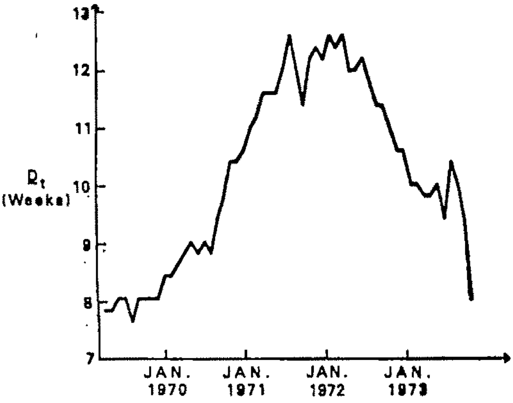


FIGURE 1

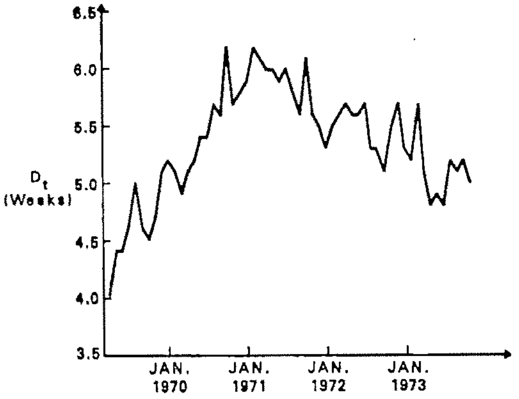


FIGURE 2

$$v_t P_t = (n_t q/k)(1 - j(G^{ab}_{t-T}, G^{ab}_t, G^{bc}_{t+T})) \tag{13}$$

From April 1969 to the end of 1973, data on the seasonally adjusted number of vacancies (in thousands) in *U.S.* manufacturing, v^m_t, have been published by the *BLS*. Figure 3 charts fluctuations in v^m_t for this period. Assume the number of vacancies for the entire economy v_t is some multiple M of v^m_t. Then an expression for P_t is obtained by dividing equation (13) by Mv^m_t. Specifically,

$$P_t = (n_t q/kM) \cdot (1 - j(G^{ab}_{t-T}, G^{ab}_t, G^{bc}_{t+T}))/v^m_t \tag{14}$$

In order to compute a measure of P_t over time, substitute into equation (14) the values of $(1-j(G^{ab}_{t-T}, G^{ab}_t, G^{bc}_{t+T}))$ previously estimated using equation (12), and the values of v^m_t. It was arbitrarily assumed that $n_t q/kM$ equaled two hundred, resulting in arbitrary absolute levels of P_t. However, a different figure for $n_t q/kM$ for the entire period will not alter the observed relative changes in P_t over time which are the data desired. Figure 4 presents the estimated values of P_t for the period April 1969 to November 1973.

A strong assumption in the above analysis is that $n_t q/kM$ is constant over time. The assumed constancy does not mean that in fact n_t (the number of firms), q (the number of occupations), k (a measure of search efficiency), and M (the ratio of the number of vacancies in the economy to the

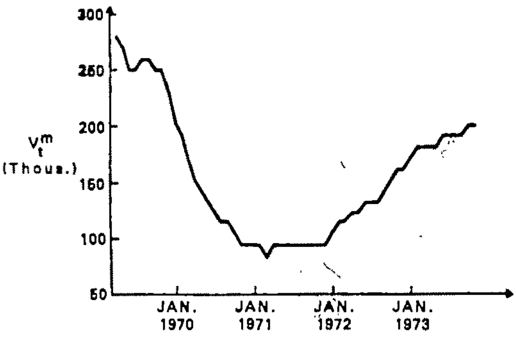


FIGURE 3

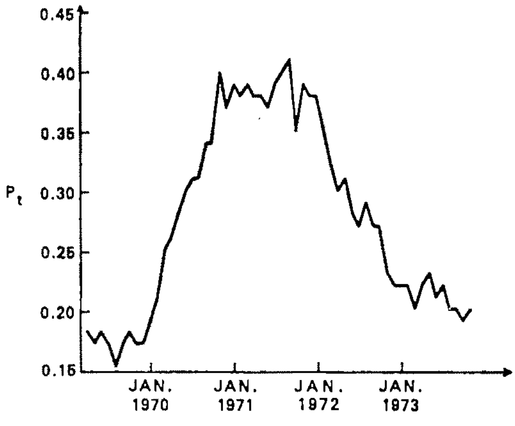


FIGURE 4

number in manufacturing) may not be cyclically sensitive. Rather it reflects the proposition that the net effect of possible cyclical changes in these variables on $n_t q/kM$, though it may bias the reported estimates of P_t, will not alter the observed cyclical pattern of change in P_t.

III. Interpretation of the Analysis and Concluding Remarks

Figures 3 and 4 presented above indicate that a decline in the number of vacancies, by decreasing the expected return to search, increases an individual's probability of accepting a wage offer P_t. This suggests that regarding P_t, Effect 2 (a reduced expected return to search) dominates Effect 3 (an unanticipated fall in the wage offer distribution). Further, Figures 2, 3, and 4 indicate that the average duration of unemployment D_t is inversely related to the number of vacancies but directly related to P_t. These relations suggest that regarding D_t, Effect 1 dominates the net result of Effects 2 and 3 on P_t. This observation also tells us the net effect that a decrease in the frequency of job offers and the resulting increase in P_t has on the unemployment rate. In fact, by equation (7), the unemployment rate rises.[12]

It is of interest to note that for the entire period, P_t seemed to be influenced mainly by the availability of jobs. This result conflicts with the view that workers, specifically unemployed job searchers, react primarily to changes in the actual wage offer distribution relative to their expectations concerning wage offers. For example, according to Mortensen, a falling unemployment rate is due to, in part, " . . . an increase in the actual rate of wage inflation . . . relative to the expected . . . (which) increases the acceptance rate (P_t) because all offers appear to have increased relative to the market average" (1970b, p. 206). Mortensen states that, though temporary, this effect is important. In fact, he suggests that this lag in the adjustment of the rate of expected wage inflation to the actual rate provides an explanation of the short-run Phillips curve. The evidence presented in this paper, however, suggests that a falling unemployment rate coincides with a falling P_t. Thus our analysis casts further " . . . doubt on [those] mechanisms proposed [by, among others, Mortensen] for reconciling Phillips curve phenomena in a price theory framework" (to quote Parsons, p. 398).[13]

To conclude, several qualifications should be cited. First, if one drops the assumption of not more than one vacancy in a particular occupation existing at any firm, this would mean that a halving of the number of vacancies may not double the time required to find a vacancy. Thus, in equation (14) the number of vacancies may enter in the form $(v_t^m)^\beta$ where $\beta < 1$. Secondly, it is likely that the ease of locating a vacancy (i.e., receiving a wage offer) will depend inversely on the number of job searchers. Assuming the number unemployed varies inversely with the number of vacancies, this suggests that in equation (14) the number of vacancies enter in the form $(v_t^m)^\beta$ where $\beta > 1$. The model, by abstracting from the above phenomena, makes the assumption that their net influence will not alter the results substantially.

---

[12] This analysis abstracts from other factors affecting the steady-state unemployment rate which may appear through changes in the variable F_t in equation (7).

[13] Parsons, in examining quit rates, found that if the vacancy rate and relative wages were held fixed then changes in a worker's own wage would not have any systematic effect on quit behavior. Restating, Parsons' evidence concerning quit rates suggests ". . . that wage 'illusion' is not an extremely strong effect" (p. 400). The results of this paper suggest that the unemployed job searcher is similarly not strongly afflicted with wage illusion in determining whether to accept or reject a wage offer. Further, our results are more appropriate for questioning theories of the Phillips curve such as Mortensen's than are Parsons' results. (See Parsons, fn. 23, for more on this.)

APPENDIX

Following is a derivation of the expression for the change in the average duration of unemployment D_t due to a change in the number of vacancies v_t. Assume that $f(w_t)$, the actual distribution of wage offers, equals $f^e(w_t)$, the perceived distribution. Given

$$\text{(A1)} \qquad D_t = 1/\theta_t P_t = \left(\frac{n_t q}{v_t k}\right)\left(\frac{1}{P_t}\right)$$

Then

$$\text{(A2)} \qquad \partial D_t/\partial v_t = -\left(\frac{n_t q}{v^2 k}\right)\left(\frac{1}{P_t}\right) - \left(\frac{n_t q}{v_t k}\right)\left(\frac{\partial P_t/\partial v_t}{P^2}\right)$$

where

$$\text{(A3)} \qquad \partial P_t/\partial v_t = (\partial P_t/\partial a_t)(\partial a_t/\partial v_t)$$

The exact form of equation (A3) depends on whether equation (3) or (4) is considered. From equation (3) we have

$$\text{(A4)} \qquad \partial P_t/\partial v_t = -f(a_t)\,\frac{n_t q}{v^2 k}\,(C)/HP_t$$

Substituting (A4) into (A2), equation (A2) becomes

$$\text{(A5)} \qquad \partial D_t/\partial v_t = (D_t)\left(\frac{1}{v_t}\right)\left(-1 + \frac{f(a_t)(C)D_t}{P_t H}\right)$$

Using equation (4) instead, equation (A3) may be expressed as

$$\text{(A6)} \qquad \partial P_t/\partial v_t = -f(a_t)\,\frac{n_t q}{v_t^2 k}\,(C + a_t) \Bigg/ \left(\frac{P_t}{s} + \frac{1}{\theta_t}\right)$$

Substituting (A6) into (A2),

$$\text{(A7)} \qquad \partial D_t/\partial v_t = \frac{D_t}{v_t}\left(-1 + \frac{f(a_t)(C + a_t)D_t}{P_t/s + 1/\theta_t}\right)$$

Equation (A5) or (A7) may theoretically be either positive or negative in sign. However, empirical evidence suggests that $\partial D_t/\partial v_t < 0$.

Equation (A5) or (A7) may be rearranged to obtain an expression for the vacancy elasticity of the duration of unemployment, denoted E_D. Specifically, for equation (A5)

$$\text{(A8)} \qquad E_D = -\frac{\partial D_t}{\partial v_t}\frac{v_t}{D_t} = \left(1 - \frac{f(a_t)(C)D_t}{P_t H}\right)$$

and for equation (A7)

$$\text{(A9)} \qquad E_D = -\frac{\partial D_t}{\partial v_t}\frac{v_t}{D_t} = \left(1 - \frac{f(a_t)(C + a_t)D_t}{P_t/s + 1/Q_t}\right).$$

Equation (A8) or (A9) and the evidence presented in this paper suggest that $0 < E_D < 1$ (inelastic). The responsiveness of the duration of unemployment to changes in the number of vacancies will depend on the parameters in equation (A8) or (A9) which include among other things, the particular wage offer distribution, the expected length of employment H, and the monetary cost of continued search C.

REFERENCES

R. J. Barro and H. I. Grossman, *Money, Employment, and Inflation,* New York 1975.

K. Burdett, "Theories of Search in a Labor Market," tech. anal. pap. no. 13, U.S. Dept. of Labor, Washington 1973.

J. P. Danforth, "Expected Utility, Mandatory Retirement and Job Search," Center for Econ. Res. disc. pap. no. 74–41, Univ. Minnesota, June 1974.

M. H. DeGroot, "Some Problems of Optimal Stopping," *J. Royal Statist. Soc.,* Mar. 1968, *30,* 108–22.

———, *Optimal Statistical Decisions,* New York 1970.

R. Flanagan, "The U.S. Phillips Curve and International Unemployment Rate Differentials," *Amer. Econ. Rev.,* Mar. 1973, *63,* 117–31.

R. Gronau, "Information and Frictional Unemployment," *Amer. Econ. Rev.*, June 1971, *61*, 290–301.

C. C. Holt, "How Can the Phillips Curve Be Moved to Reduce Both Inflation and Unemployment?," in E. S. Phelps et al., eds., *The Microeconomic Foundations of Employment and Inflation Theory*, New York 1970.

———, "Segmentation of the Labor Market: Rejoinder," *Amer. Econ. Rev.*, June 1975, *65*, 491.

H. B. Kaitz, "Analyzing the Lengths of Spells of Unemployment," *Mon. Lab. Rev.*, Nov. 1970, *93*, 11–20.

R. E. Lucas, Jr. and E. C. Prescott, "Equilibrium Search and Unemployment," mimeo. 1973.

J. J. McCall, "Economics of Information and Job Search," *Quart. J. Econ.*, Feb. 1970, *84*, 113–26.

D. T. Mortensen, (1970a) "Job Search, the Duration of Unemployment, and the Phillips Curve," *Amer. Econ. Rev.*, Dec. 1970, *60*, 847–62.

———, (1970b) "A Theory of Wage and Employment Dynamics," in E. S. Phelps et al., eds., *The Microeconomic Foundations of Employment and Inflation Theory*, New York 1970.

———, "Job Matching Under Imperfect Information," paper presented at Conference on Evaluating the Labor Market Effects of Social Programs, Princeton 1974.

D. O. Parsons, "Quit Rates Over Time: A Search and Information Approach," *Amer. Econ. Rev.*, June 1973, *63*, 390–401.

M. Rothschild, "Models of Market Organization with Imperfect Information: A Survey," *J. Polit. Econ.*, Nov./Dec. 1973, *81*, 1283–1308.

———, "Searching for the Lowest Price When the Distribution of Prices is Unknown," *J. Polit. Econ.*, July/Aug. 1974, *82*, 689–711.

S. Salant, "A Theory of Sort," paper presented at meetings of Amer. Econ. Assn., New York, Dec. 1973.

S. C. Salop, "Systematic Job Search and Unemployment," *Rev. Econ. Stud.*, Apr. 1973, *40*, 191–201.

J. Seater, "A Unified Model of Consumption, Labor Supply and Job Search," unpublished doctoral dissertation, Brown Univ. 1974.

U.S. Bureau of Labor Statistics (*BLS*), *Employment and Earnings*, various issues, Washington 1969–74.

Private Preference for Collective Goods Revealed Through Voting on Referenda

By ROBERT DEACON AND PERRY SHAPIRO*

When studying consumer preferences or demands for private commodities, researchers have the advantage (at least in principle) of observing the results of individual optimizing decisions in the market. This, of course, is not the case for collectively supplied goods and services. Since the level at which such goods are provided is a collective decision, the observation that an individual with a given income consumes a particular bundle of collective goods is not evidence that this bundle is preferred to all others that are not more expensive.

Empirical studies of the demand for public goods[1] have, in general, examined the relationship between jurisdictional expenditures and various socioeconomic aspects of the constituent populations.[2] These studies typically rely upon a simple concept of political equilibrium in which the actions of a political jurisdiction reflect the desires of the median voter. Utilizing data on median levels of various population characteristics, certain empirical relationships between these characteristics and public expenditures are derived. If the assumptions on the political process are satisfied, these relationships are indicative of median voters' demands.

The problem of estimating individual demands for collective goods may be approached in a second way, by examining voting behavior on referenda related specifically to the provision of public goods.[3] This is the approach taken in the present paper. The presentation begins by developing a model of the individual's voting decision. The problem of using aggregate responses to draw inferences on individual preferences is then addressed and some operational assumptions are employed. Next, the technique of conditional logit analysis is adapted to analyze the categorical voting responses, "yes," "no," and "abstain." Two recent California referenda are used to examine the practical possibility of using voting data to estimate demands for collective goods.

I. The Individual Voting Decision

Under any public policy k an individual i is constrained to consume a vector of collective goods q provided at a collectively chosen level q_k, and faces a total tax liability S_k^i (derived from a system of predetermined tax rates or prices) for the provision of these goods. Within the con-

---

* University of California, Santa Barbara. This research was supported by a grant obtained through the Marine Science Institute, University of California, Santa Barbara. We wish to acknowledge research assistance from Jeff Newton and helpful comments from Anthony Barkume and an anonymous referee, while retaining full responsibility for any errors which may remain.

[1] The terms "public goods" and "collective goods" are used interchangeably to denote goods and services provided by collective decisions, and do not necessarily imply indivisibilities in consumption.

[2] James Barr and Otto Davis, Thomas Borcherding and Robert Deacon, and Theodore Bergstrom and Robert Goodman provide examples of this approach.

[3] See, for example, Edward Banfield and James Wilson, William Birdsall, and Shapiro.

text of this public policy, the consumer chooses a vector of private commodities $\mathbf{x}^i$ subject to a budget constraint.

$$(1) \qquad (p_k)(\mathbf{x}^i) = I_k^i - S_k^i$$

where p_k is a vector of private goods prices (assumed to be the same for all individuals) and I_k^i is money income. Prices and money income appear with policy subscripts to allow for the very real possibility that they may be influenced, directly or indirectly, by public policy.

Each individual is assumed to have a complete semiorder (see R. Duncan Luce) over all possible bundles of private and collective goods. The concept of a semiorder, while similar to the conventionally assumed preorder, allows for intransitivity in the indifference relation (the reason for introducing such intransitivity becomes apparent below). If P and I denote respectively the relations preferred and indifferent, and if a, b, c are commodity bundles, a semiorder allows for the possibility that aIb, bIc, but aPc. A natural interpretation of this is the notion of threshold levels in individual perception. Luce has shown that the semiorder may be used to induce a weak order on the consumption set and one may then postulate the existence of a utility function U to represent the weak order. While the utility function preserves the weak order, it cannot alone represent the concept of indifference. To capture indifference it is necessary to further define functions δ such that aIb if and only if $-\delta \leq U(a) - U(b) \leq \delta$, where δ takes on the role of a threshold level for utility discrimination. The function δ is defined only in terms of a given utility function so that any monotone increasing transformation of U would require a commensurate transformation of δ to preserve the semiorder.[4]

The individual consumer-voter is assumed to possess a differentiable quasiconcave utility function

$$(2) \qquad U^i = U^i(\mathbf{x}^i, q_k)$$

with vectors of private and collective goods as arguments. Although elements of imprecision have been introduced into individual preferences, it is possible to define the highest level of utility attainable by the individual subject to the budget constraint. Assuming the budget constraint to be satisfied exactly, the solution may be used to solve for the vector of private consumptions in terms of prices, disposable income, and predetermined levels of collective consumption. The highest attainable utility may then be written in indirect form as

$$(3) \qquad \max_{\mathbf{x}^i} U^i(\mathbf{x}^i, q_k) = V^i(q_k, p_k, I_k^i - S_k^i)$$

a function of variables each of which may be affected by the choice of public policy.

Analysis of the individual's role in the collective decision-making process is confined to the act of voting on public referenda. Any proposed change in public policy represents a potential change in one or more of the parameters q_k, p_k, $I_k^i - S_k^i$, which in turn implies a change in V^i. Denoting any two alternatives offered on a ballot by $k=0, 1$, highest attainable utilities under alternative outcomes are

$$(4) \qquad V^i(q_0, p_0, I_0^i - S_0^i) = V_0^i$$

$$(5) \qquad V^i(q_1, p_1, I_1^i - S_1^i) = V_1^i$$

Voters are assumed to cast ballots in their perceived self-interest, i.e., the voting decision is reached by comparing highest attainable utility under the outcomes of alternative policies.

---

[4] Luce defines the "just noticeable differences" δ with respect to individual commodity bundles and allows for the possibility that there may be distinct upper and lower just noticeable differences for any particular commodity bundle. To ease exposition we have assumed that upper and lower just noticeable differences are equal. One unfortunate consequence of this assumption is that it will not, in general, be maintained under monotonic transformation of the utility function. We are indebted to John Howell for this point.

While this specification captures the essence of individual preferences for alternative policy outcomes, the decision to go to the polls at all has been ignored. Most elections offer choices on a variety of issues (public referenda, candidates for political office, etc.). In a general model of voting, the individual would weigh the expected benefit from voting on a range of issues with the costs of going to the polls (time, transportation, etc.), and would not vote at all unless the former exceeded the latter (see Anthony Downs). The aim here is more modest in that the individual's presence at the polls is taken as datum and the decision to vote on a marginal issue is analyzed. If the outcomes of alternative policy proposals were known with certainty and potential utilities could be weighed with knife-edged precision, a simple decision rule would emerge; individual i would vote for alternative 1 if and only if $V_1^i > V_0^i$. Once at the polls the cost of voting on a marginal issue is virtually zero and abstention on a particular issue would occur only in cases of perfect indifference, $V_1^i = V_0^i$. It follows that very few voters would be expected to abstain on any ballot proposition.

In fact, however, one typically observes that a substantial portion of all citizens at the polls will not vote at all on a number of issues. A natural explanation for abstention in the voting booth is provided by the concept of a semiorder and its representation in thresholds for utility discrimination. The decision rule which follows is that the individual will prefer and vote for alternative 1 if highest attainable utility under this outcome is perceptibly greater than attainable utility under alternative 0.[5] If the difference in utility under alternative outcomes is not perceptible, the individual will be indifferent. The following analysis is developed in terms of differences in potential utility under alternative outcomes, which for convenience is written

$$(6)\quad V_1^i - V_0^i = \Delta V^i(q_0, p_0, I_0^i - S_0^i, \Delta q, \Delta p, \Delta I^i - \Delta S^i)$$

where $\Delta q = q_1 - q_0$, etc. Letting x represent the vector of arguments in ΔV^i, (6) may be expressed more compactly as

$$(6')\qquad V_1^i - V_0^i = \Delta V^i(x^i)$$

Designating a vote for 1 as a yes vote, the decision rule becomes: vote yes if $\Delta V^i > \delta^i$; vote no if $\Delta V^i < -\delta^i$; abstain only if $|\Delta V^i| \leq \delta^i$.

While this specification allows for diversity in individual preferences, it does so only through the form of the functions V^i. An operational way to allow for differences in tastes is to assume that such variations are reflected in certain socioeconomic characteristics of the individual. The differences in potential utility then become

$$(7)\qquad \Delta V^i(x^i) = \Delta V(x^i, y^i)$$

where y^i is a vector of individual socioeconomic characteristics.

The term δ^i can be interpreted as a measure of the precision with which an individual is able to distinguish between alternatives. It seems likely that the degree of precision would be related to such factors as the individual's experience, predisposition, and contextual knowledge. Education, for example, might enhance one's ability to discern preference among alternatives, in which case δ^i would be inversely related to educational attainment. To allow for the possibility that these perception thresholds may be influenced by education as well as other socioeconomic characteristics, the specification

---

[5] A variety of alternative reasons for abstention at the polls could be posited. One possibility arises from uncertainty or lack of information on the potential outcomes of policy proposals. Although such uncertainty surely exists, it would not necessarily induce the individual to abstain from making a choice. With most referenda, one potential outcome represents the status quo (e.g., a no vote) which is, presumably, known with relative precision. In such cases, normal risk aversion would not tend to favor abstention, but would, rather, bias the voting choice toward the status quo.

$$\delta^i = \delta(y^i) \tag{8}$$

is adopted.[6]

The model was developed to suggest variables which might influence individual voting decisions, and to facilitate the interpretation of empirical results. It would be impossible (or at least undesirable) to incorporate into the empirical analysis all variables suggested by the model. However, by examining a specific ballot proposal in light of these expressions one may select a subset of the variables which, a priori, appear to be important. For example, a proposed policy change may affect only the jth public good, one private price, and the levels of income of a certain class of citizens. If, further, all citizens initially consume public goods at the same level and face the same private prices, the model can be simplified considerably for purposes of estimation.

II. An Econometric Specification

Since individual ballot choices are not disclosed, only the aggregate responses of groups of individuals are observed. Thus it is necessary to employ several assumptions on the patterns of preferences in voting groups in order to make the model operational. As a first step, let $\bar{z}_m$ and $\bar{y}_m$ denote the vectors of mean attributes and socioeconomic characteristics of members of voting group m. Within the group, assume that ΔV^i is a random variable distributed in a known fashion with mean $\mu(\bar{z}_m, \bar{y}_m)$ and variance σ^2. The distribution of ΔV^i for jurisdiction m is written $g(\Delta V^i | \bar{z}_m, \bar{y}_m)$ and is, therefore, conditional on the vectors of mean attributes.[7] While the means of the distributions will vary across voting groups as mean attributes differ, all distributions are assumed to possess a common variance. Within any voting group characterized by the vectors $\bar{z}_m, \bar{y}_m$, the probability of drawing at random an individual who will vote no (vote for alternative 0) is defined to be

$$P(N | \bar{z}_m, \bar{y}_m) = \int_{-\infty}^{-\delta_m} g(\Delta V^i | \bar{z}_m, \bar{y}_m) d(\Delta V^i) \tag{9}$$

where δ_m is, as yet, unspecified. The probability of drawing a yes vote is

$$P(Y | \bar{z}_m, \bar{y}_m) = \int_{\delta_m}^{\infty} g(\Delta V^i | \bar{z}_m, \bar{y}_m) d(\Delta V^i) \tag{10}$$

and the probability of finding an abstainer is

$$1 - P(N | \bar{z}_m, \bar{y}_m) - P(Y | \bar{z}_m, \bar{y}_m) \tag{11}$$

Although the notation employed is suggestive, the interpretation of δ_m in (9) and (10) is not unambiguous. It has meaning in terms of the individual choice model in Section I only if one adopts the highly restrictive assumption that individual δ^i are identical for all voters in a given jurisdiction m and are a function of the mean socioeconomic characteristics of that group, i.e., $\delta^i = \delta(\bar{y}_m)$. Rather than adopt this

---

[6] Elements of the vector x^i are not included in (8) as they are related to levels of utility, but not to the form of the function.

[7] The reader may feel uneasy with the apparent cardinality in this formulation. Such objections could have been avoided by specifying that individuals impute dollar values to potential policy changes, by defining $R^i = (V_1^i - V_0^i)/\lambda^i$ where λ^i is the individual's marginal utility of income (assumed to be constant with respect to variables influenced by the policy). It can be shown that a first-order approximation to R^i is given by

$$R^i = \sum_j M_j^i \Delta q_j + (\Delta I^i - \Delta S^i - \sum_k x_k^i \Delta p_k)$$

where M_j^i is the individual's marginal rate of substitution between q_j and disposable income. One could then specify a distribution for the R^i. This transformation was not undertaken in the text as it would complicate the presentation with little effect upon the development of the model or interpretation of results.

specification, we merely employ the operating assumption that δ_m is a function of $\bar{y}_m$, and do not directly relate δ_m to individual behavior.

The final tally from jurisdiction m in which T_m citizens voted can be viewed as the outcome of a trinomial experiment performed on a sample of size T_m of the population in the jurisdiction. The observed sample proportions $\hat{P}$ (percent who voted yes or no) are therefore equal to the true probabilities in (9) and (10) plus a random error term. The density functions g in (9) and (10) are taken to be logistic, so that the observed sample proportions are

$$\hat{P}(N_m) = [1+\exp((\delta_m+\mu_m)/\tau)]^{-1} + \epsilon_m^N \tag{12}$$

$$\hat{P}(Y_m) = [1+\exp((\delta_m-\mu_m)/\tau)]^{-1} + \epsilon_m^Y \tag{13}$$

where $\mu_m \equiv \mu(\bar{z}_m, \bar{y}_m)$, $\tau = \sigma\sqrt{3}/\pi$

and the error terms follow a multinomial distribution.

Given the assumed logistic distribution, it is convenient to work with the "logit" transformation

$$\begin{aligned} (12')\quad l(N_m) &\equiv ln\,[\hat{P}(N_m)/(1-\hat{P}(N_m))] \\ &= -(\mu_m+\delta_m)/\tau + \nu_m^N \end{aligned}$$

$$\begin{aligned} (13')\quad l(Y_m) &\equiv ln\,[\hat{P}(Y_m)/(1-\hat{P}(Y_m))] \\ &= (\mu_m-\delta_m)/\tau + \nu_m^Y \end{aligned}$$

where l denotes the logit operator, ln is the natural logarithm, and ν_m^N and ν_m^Y are the transformed error terms.

An estimate of δ_m may be obtained directly by forming the sum

$$l(Y_m) + l(N_m) = -2\delta_m/\tau + \nu_m^R \tag{14}$$

where $\nu_M^R = \nu_m^Y + \nu_m^N$.

For purposes of estimation, the functions $\mu(\cdot)$ and $\delta(\cdot)$ are taken to be linear so that

$$\mu_m/\tau = \bar{z}_m\alpha' + \bar{y}_m\beta' \tag{15}$$

and $\delta_m/\tau = \bar{y}_m\gamma'$

where α, β, and γ are row vectors of constants. Equations (13'), (14), and (15) are combined to yield the following estimating equations

$$l(Y_m) = \bar{z}_m\alpha' + \bar{y}_m(\beta'-\gamma') + \nu_m^Y \tag{16}$$

$$l(Y_m) + l(N_m) = -2\bar{y}_m\gamma' + \nu_m^R \tag{17}$$

The parameters of (16) and (17) were jointly estimated using the minimum logit χ^2 approach (the details of this technique are outlined in J. E. Grizzle et al.).

III. Estimation

While this model provides a concise framework for analysis, the complexities of actual political referenda and data limitations often necessitate the use of less than ideal explanatory variables in empirical analysis. For example, the potential difference in disposable income under alternative outcomes ($\Delta I^i - \Delta S^i$ in (6)) cannot be observed directly. However, it is often reasonable to hypothesize that the incomes of certain segments of the populace (occupational groups, for instance) would be significantly affected by a particular policy. To capture such effects, the proportion of voters in an affected group may be included among the set of explanatory variables. Another type of variable which naturally poses problems for measurement is the level of consumption of various collective goods (q_0 and Δq in (6)). Because these variables are at best difficult to quantify, dummy variables may be used to partition the set of voting aggregates into groups within which collective consumption levels appear to be uniform. The regression coefficients then provide the estimates sought. All of these considerations are explained below in the context of specific referenda.

Two referenda which appeared on recent statewide ballots in California were chosen for analysis. They dealt with conservation of the California coastline and provision of public financing for rapid transit. Both measures elicited a great deal of voter interest and each was concerned explicitly with provision of collective goods. Furthermore, both related to environmental protection.

A. *The Coastal Zone Conservation Act*

The Coastal Zone Conservation Act was submitted to the voters (and adopted) in November 1972. It was an initiative measure designed to create six regional commissions and one statewide commission to oversee development in the California coastal zone. Although the stated purpose of the act was to develop a comprehensive plan for coastal zone land use, the major issues in the campaign hinged upon the power of the regional commissions to prohibit development within the "permit area," a narrow corridor within the coastal zone. The act proposed to preserve the natural coastal environment by halting or delaying development for a period of three years, after which the land use plan would be submitted to the voters for approval. Successful implementation of the act would lead to limiting the rights of some land owners to develop their property. The act was also expected to adversely affect incomes and employment in the construction industry, to delay the siting and construction of new power generating plants, and to have differential impacts upon property values in the affected areas. The measure appropriated a total of $5 million from state funds for three years of operation so that the direct tax burden of the act (exclusive of any effect which changing property values may exert upon local tax bases) was negligible.

The voting aggregates used in the empirical analysis were 334 California cities. Information reported by the California Secretary of State was used to calculate the percentage (as a function of total votes T cast for all presidential candidates), of affirmative Y and negative N votes on Proposition 20 in various cities. Data from the same document were used to compute F, the fraction of votes cast for the Republican or American Independent presidential candidate. Observations on per capita income I, percent of work force employed as laborers or construction craftsmen W, and the mean number of school years completed E in each city were taken from the 1970 U.S. Census. In addition, two partitions of the state were formed. The boundaries of the first partition, $A_1 \ldots A_6$, were drawn so that travel time to the coast was roughly equal from all points within a given zone. The second partition, Q_1, Q_2, and Q_3 divided the state into southern, central, and northern regions, respectively. Dummy variables were defined to identify each city's location in these partitions.

The reason for including income is obvious and, unless the coastal environment is an inferior good, one would expect its coefficient to be positive. The variable W corresponds to ΔI in equation (6) and was included to reflect the potential impact which the policy was expected to have upon incomes in the construction industry: a negative coefficient is expected. The variables E and F correspond to elements of $\bar{y}$ in equations (16) and (17) and were incorporated to capture variations in preference patterns among voters. The dummy variables Q_1, Q_2, Q_3 were employed to reflect differences in the initial condition of the environment. As one moves from south to north (Q_1 to Q_3), the initial quality of the environment becomes higher and, on the basis of a simple diminishing marginal utility argument, one would expect the magnitudes of the coefficients to become progressively smaller. The access zones $A_1 \ldots A_6$ were used to capture differences in the expected change in the level of the collective good (Δq in equation (6)). It was assumed that the expected change in environmental quality was equal for all citizens in the coastal zone (A_1).

However, at locations further inland ($A_2 \ldots A_5$ progressively), individual access to the coastline declines so that any given change in the coastal environment represents a smaller change in the level of environmental quality experienced by the individual.

In terms of the model it was assumed that initial price levels (p_0) and any changes in prices (Δp) expected to follow from the proposed policy (for example, higher future prices for electric power due to delays in expanding generating capacity) were uniform throughout the state. Any price effects would therefore be subsumed in the constant term.

Of all the impacts the policy was expected to have, perhaps the most difficult to sort out empirically involve differential changes in land values. Presumably, owners of undeveloped land in the permit area were faced with a capital loss if the measure was adopted. On the other hand, owners of developed property in the permit area probably stood to gain on balance from the initiative. The data lacked sufficient detail to separate residents of the permit area from those of the coastal zone (i.e., access zone A_1 included cities in both the permit area and the coastal zone). For these reasons, the expected sign of A_1 is ambiguous as it will reflect a combination of accessability and property value effects.

All of the independent variables described above were incorporated in a regression equation to explain $l(Y)$ (see (16)). The socioeconomic characteristics E and F as well as per capita income I, were used as independent variables in estimating (17). Application of the joint estimation technique to this system of equations yielded the estimates displayed in Table 1. Regression coefficients are shown in the first two columns and are interpreted as the overall effects of the independent variables upon the probability of voting in favor of the proposition (col. 1) and the probability of voting on this initiative at all (col. 2). Estimates of α, β, and γ, the items of primary interest, were obtained by linear transformation of the regression coefficients (see equations (16) and (17)) and are reported in columns three and four. These are interpreted as the effects of the independent variables upon expected changes in utility (col. 3) and upon the perception thresholds (col. 4). It is the items in these last two columns which are the subject of discussion below.

In general, the parameter estimates have the expected signs. The income coefficient suggests that the collective good, the natural coastal environment, is normal although the effect is not highly significant. There are, however, considerations which render the interpretation of this item somewhat ambiguous. The nature of the initiative indicated that it would significantly affect property values, particularly in the coastal zone and permit area. The data available did not allow these property value effects to be represented directly and to the extent that property values and incomes are correlated, the income term may be capturing some of these influences. The negative and highly significant coefficient associated with percent laborers corroborates the hypothesis that the initiative was expected to adversely affect incomes and employment in that sector.

The coefficients of locational zones Q_1 and Q_3 (south and north, respectively) also appear quite reasonable. As noted earlier, the initial level of environmental quality was higher in the north than in the south so that a larger coefficient (stronger support) in the south than in the north was anticipated. With regard to the access zones $A_1 \ldots A_4$, none of the coefficients obtained are significantly different from zero (at 5 percent). Although the interpretation of A_1 is unclear for reasons described above, it was expected that the

TABLE 1—COASTAL ZONE CONSERVATION ACT[a]

Independent Variables	Regression Coefficients: $l(Y)$ $(\alpha, \beta-\gamma)$	Regression Coefficients: $l(Y)+l(N)$ (-2γ)	Parameter Estimates: α, β	Parameter Estimates: γ
Constant	0.0985	−0.4703	0.0985	0.2351
	(0.3128)	(0.0229)	(0.3128)	(0.0114)
I (income)	0.0162	0.0174	0.0074	−0.0087
	(0.0165)	(0.0035)	(0.0165)	(0.0018)
W (percent laborers)	−8.9952		−8.9952	
	(1.2218)		(1.2218)	
E (education)	0.1120	0.0100	0.1070	−0.0050
	(0.0177)	(0.0012)	(0.0165)	(0.0006)
F (political preference)	−1.6793	0.2005	−1.7795	−0.1002
	(0.1074)	(0.409)	(0.1058)	(0.0204)
Q_1 (southern)	−0.0545		−0.0545	
	(0.0268)		(0.0268)	
Q_3 (northern)	−0.1630		−0.1630	
	(0.0968)		(0.0968)	
A_1 (access zone 1)	−0.1294		−0.1294	
	(0.1965)		(0.1965)	
A_2 (access zone 2)	−0.0521		−0.0521	
	(0.1962)		(0.1962)	
A_3 (access zone 3)	0.0645		0.0645	
	(0.2019)		(0.2019)	
A_4 (access zone 4)	−0.1923		−0.1923	
	(0.1981)		(0.1981)	

Sources: Dependent variables and F were formed from data reported in California Secretary of State, *Statement of the Vote*. Variables I, W, and E were obtained from U.S. Bureau of the Census. A description of the dummy variables Q_1, Q_3 and $A_1 \ldots A_4$ is provided in the text.

[a] Standard errors in parentheses.

coefficients of A_2, A_3, A_4 would form a monotonic sequence, as these zones represent progressively smaller changes in the level of the collective good. No such pattern is, however, evident.

The variables E (education) and F (political preference), included to reflect differences in tastes, are highly significant in both equations. Since the issue proposed to usurp some of the traditional rights of property owners, the strong disapproval by voters with relatively conservative political views (as evidenced by the negative coefficient on F in the third column of Table 1) was anticipated. Further, the coefficient of education in that equation lends support to the popular belief that environmental measures are supported by and benefit the better educated segments of society.

Figures in the fourth column of Table 1 relate to the perception thresholds δ. The negative and significant coefficient associated with education is quite plausible as it suggests that the size of perception thresholds (the degree of imprecision in perceiving preference) is inversely related to educational attainment.[8] No a priori hypotheses were formed on the role of political preference and income in the second equation although both terms appear with significant coefficients.

Our inability to measure in any quantitative fashion several of the variables deemed important determinants of the voting choice led to use of proxies for these unquantified influences (for example, the dummy variables $A_1 \dots A_5, Q_1 \dots Q_3$, as well as E and F). To examine the significance of these sets of variables taken as groups, an asymptotic F-statistic was used.[9] The first null hypothesis examined was that the set of coefficients of access zone variables, $A_1 \dots A_4$ is not significantly different from zero. This hypothesis was rejected at 1 percent significance ($F(4,653)=5.7$). The null hypothesis that the coefficients of locational zones Q_1 and Q_3 are not significant, produced an F-statistic of $F(2,653)=3.0$ which coincides with the rejection level at 5 percent significance. The hypotheses that the coefficients of the preference variables E and F are not significantly different from zero in the two equations were clearly rejected at 1 percent ($F(2,653)=140.8$ in the first equation and $F(2,653)=72.1$ in the second equation). General comments and conclusions are reserved for a later section.

---

[8] The reader should keep in mind the previously discussed difficulties associated with interpreting this term.

B. *The Rapid Transit Initiative*

The second issue dealt with public financing of mass transit. By California law, the revenue collected by the state from the tax on gasoline may be used only for public highway purposes. The revenue so collected may be used for the construction and maintenance of state and local roads as well as related administrative and law enforcement activities. Proposition 18, which appeared on the November 1970 ballot (and was rejected by the voters), proposed to amend this law by allowing counties and cities to use up to 25 percent of their share of gas tax revenues for the development of public mass transportation. These revenues could be so used only if specific authorization was approved by a majority of the affected voters. The funds could be used not only to construct and operate new mass transit systems but also to meet interest payments or retire any bonded debt incurred in the construction of such a system. Furthermore, the proposed amendment allowed the state, through legislative fiat, to use an unlimited portion of highway user tax revenues for the control of environmental pollution caused by motor vehicles.

As with Proposition 20, this proposed amendment was viewed in part as an environmental protection measure. By discouraging the use of automobiles and by fostering pollution control research, Proposition 18 intended to help the state ". . . deal with the increasingly grave problem of air pollution"[10] (capitals in the original). In addition to its environmental benefits, the measure was viewed as a way to diminish traffic congestion on the existing freeway system. It was argued that "Each new freeway is loaded to capacity in our urban areas the day it opens, and all the rest of the system becomes increasingly overloaded."[11]

The proposed amendment had one additional feature that was not mentioned by its proponents. If adopted, auto tax revenues could be used not only to finance new construction, but also to retire bonded debt on existing systems. This feature of the amendment made it particularly appealing to the member communities of the San Francisco Bay Area Rapid Transit District (*BART*). The *BART* fixed rail system was considerably more costly than its proponents had originally anticipated, and the system found itself with an unexpectedly large bonded debt. Provisions of Proposition 18 would have allowed *BART* to service its debt without finding new tax revenues.

---

[9] The F-statistic was derived from statistics which are only asymptotically χ^2 and for that reason, their ratio may not converge to an exact F distribution.

[10] See California Secretary of State, "Proposed Amendments to Constitution," *General Election*, Tuesday, Nov. 3, 1970, p. 3.

[11] See fn. 10.

Opponents to Proposition 18 argued that the state needed all the revenues it could obtain to carry out its ambitious freeway master plan. There was a chance not only of diminished state support but also diminished federal matching fund revenues if the amendment was enacted. The opponents correctly pointed out that ". . . interest charges on the *BART* system alone. . . . could siphon off many millions of dollars annually."[12] If the highway program were to go on at its present (or moderately reduced) pace, substantial tax increases would be needed. If the program were to be reduced, it would result in a substantial diminution in highway construction activity and a loss in income and employment.

On balance, the initiative offered a rather complex package of tradeoffs to the voters. It would have given communities the option of substituting mass transit expenditures for outlays on highway construction and maintenance, a tradeoff among two collective services. Further, it would have allowed the state legislature to allocate previously earmarked highway funds to air pollution control research; hence another potential substitution among collective services. Implicit in the proposition was a third tradeoff between private and public expenditure. While Proposition 18 did not specifically provide for an increase in the state gasoline tax rate, the nature of the initiative made this outcome likely.[13]

As with the Coastal Initiative, voting results for the Rapid Transit Initiative were taken from the report of the California Secretary of State. In addition to data on per capita income, educational attainment, and political preference previously described, observations on percent transport workers W', population density D, and total land area A in various cities were taken from the 1970 U.S. Census report. Finally, a dummy variable B was included to indicate membership of a city in the *BART* District.

To the extent that the measure was expected to enhance air quality, both through a substitution from private autos to rapid transit and via increased state effort in air pollution control research, the income term will reflect citizen demands for air quality. On the other hand, the act intended to reduce highway congestion and travel times. The obvious relationship between income and the cost of time spent in transit means that the income term will reflect a combination of at least these two effects. It seemed likely that adoption of Proposition 18 would force a decline in highway construction and maintenance activity. The proportion of citizens employed as transport workers (corresponding to ΔI in (16)) was included to capture the potential adverse impact which such an outcome could have upon incomes and employment in the transport industry.[14] The special circumstances existing in the *BART* District indicated inclusion of the dummy variable B in the equation explaining yes votes. Since Proposition 18

---

[12] Secretary of State as cited in fn. 10, p. 4.

[13] Proposition 18 did not propose any specific outlays on either rapid transit or air pollution control research. Rather it granted state and local governments the opportunity to make such expenditures at the expense of outlays on highways if they so decided. Hence, in deciding how to vote, citizens had to weigh the merits of the measure from the viewpoint of a subjectively formed probability that the proposed options if granted would be exercised.

[14] Presumably, adoption of Proposition 18 would have damaged the competitive position of the highway transport sector of this industry vis-à-vis the other transport modes (rail, air, etc.). While it would therefore have been desirable to separate employment in highway transport from employment in other sectors, data were not available at this level of detail. Examination of a few large cities for which such data were available indicated that trucking services accounted for a much larger share of transport industry employment than any of the other sectors.

TABLE 2—RAPID TRANSIT INITIATIVE[a]

	Regression Coefficients		Parameter Estimates	
Independent Variables	$l(Y)$ $(\alpha, \beta-\gamma)$	$l(Y)+l(N)$ (-2γ)	α, β	γ
Constant	−2.0363	−1.3138	−2.0363	0.6569
	(0.1974)	(0.0900)	(0.1974)	(0.0450)
I (income)	0.1361	−0.0110	0.1416	0.0055
	(0.0169)	(0.0080)	(0.0169)	(0.0040)
W′ (percent transport workers)	−0.0531		−0.0531	
	(0.0821)		(0.0821)	
B (*BART* District)	0.1294		0.1294	
	(0.0411)		(0.0411)	
E (education)	0.1271	0.0514	0.1014	−0.0257
	(0.0166)	(0.0080)	(0.0160)	(0.0040)
F (political preference)	−0.9043	0.6776	−1.2431	−0.3388
	(0.1523)	(0.0700)	(0.1480)	(0.0350)
D (population density)	0.1311		0.1311	
	(0.0340)		(0.0340)	
A (land area)	0.1680		0.1680	
	(0.0872)		(0.0872)	

Sources: Variables *A* and *D* were obtained from U.S. Bureau of the Census; *B* is explained in the text. For other sources, see Table 1.

[a] Standard errors in parentheses.

offered these cities the opportunity to limit tax increases by transferring existing revenue from highway uses to finance existing debt, a positive coefficient is anticipated.

Since Proposition 18 proposed to allow municipalities to substitute against auto oriented public services toward the mass transit mode, the rate at which a municipality could potentially trade off or transform these services is relevant to the voting decision. Under the hypothesis that the mass transit alternative is potentially more efficient in large dense cities, the variables *A* and *D* were included. However, large dense cities are also areas where existing levels of air quality were lowest, so that the variables *A* and *D* may in part also reflect differential citizen demands for air pollution control.[15] Fortunately, both of these effects should work in the same direction so that positive coefficients are expected for these variables.

As with the Coastal Zone Initiative, educational attainment and political preference were included to reflect variations in citizen preferences. Further, it was assumed that initial price levels and expected changes in prices were uniform throughout the state so that any price effects would be subsumed in the constant term.

The linear equation used to estimate $l(Y)$ (see (16)) incorporated all of the independent variables defined above. As before, the analog of equation (17) utilized educational attainment *E*, political preference *F*, and per capita income *I* as regressors. Joint estimation of these equations yielded the numbers reported in Table 2. Estimates of the structural parameters α, β, and γ, displayed in the third and fourth columns, were obtained by linear transformation of the estimated regression coefficients.

The signs of the parameter estimates in Table 2 are all plausible. The positive coefficient on income suggests that the demand for air quality is income elastic, or that the measure was expected to reduce congestion and travel times, or both. Although the coefficient associated with W' (percent transport workers) is not highly significant, its negative sign lends support to the hypothesis that Proposition 18 was expected to have an adverse effect upon incomes in that sector. Further, the findings corroborate contention that member communities in the *BART* District stood to benefit from Proposition 18, since it would have allowed them the opportunity to service their debt without tax increases.

---

[15] Unfortunately, comparable data on air pollution levels were not available for most of the cities in the sample.

Population density and land area also possess anticipated signs (positive). These results are consistent with the hypotheses that (i) the potential efficiency with which highway services can be transformed into rapid transit services is higher in large dense cities, and/or (ii) the comparatively low levels of air quality in such areas gives rise *ceteris paribus* to relatively intense demands for pollution control.

Both socioeconomic variables E and F appear significant, although no a priori expectations regarding the signs of their coefficients were formed. To the extent that Proposition 18 was viewed as an environmental protection measure, the coefficient on educational attainment is consistent with results obtained from the analysis of the Coastal Zone Initiative. The significant disapproval in areas with conservative voting records may reflect a preference for private auto transport over the more public mass transit mode. Estimates obtained for δ (fourth column of Table 2) are generally consistent in sign with those reported for Proposition 20 and are given a similar interpretation. The only anomalous result, the coefficient on income, has no obvious explanation.

To examine the significance of various groups of coefficients, the hypothesis testing procedure utilized in analysis of the Coastal Initiative was employed. The null hypothesis that the entire set of regressors in both equations was not significantly different from zero was rejected at 1 percent ($F(13,658)=130.3$). The hypothesis that the two variables density and land area were not significant was also rejected at 1 percent ($F(2,658)=8.5$). Finally null hypotheses were formed to examine the significance of the preference variables E and F in the two equations. The null hypotheses that these variables were not significantly different from zero in either equation yielded F-statistics of $(2,658)=45.3$ and $F(2,658)=73.8$, respectively, and were therefore both rejected at 1 percent.

IV. Conclusions

The purpose of this study was to develop a model of individual behavior that allows one to draw inferences about preferences from voting data. The results, although pertaining to only two issues, demonstrate the feasibility of this approach. Further, the results obtained cast doubt upon the notion that individuals somehow alter their preferences (or behavior) away from selfishness and toward the social good as they leave the market and enter the polling booth. To the extent that the impact of the proposed policy changes could be specified beforehand, observed voting responses were consistent with self-interest.

The two measures, although different in most respects, both dealt with improving environmental quality. One would therefore anticipate a degree of consistency between issues from variables that reflect attitudes and underlying preferences. Education, political preference, and income were chosen to represent voter predisposition. Tables 1 and 2 show a high degree of consistency in signs of preference variable coefficients obtained for the two issues. Not only are the results consistent between issues, they also support the growing evidence that environmental protection seems to be most favored by the affluent and well-educated segments of society.

REFERENCES

E. C. Banfield and J. Q. Wilson, "Voting Behavior on Municipal Expenditures: A Study in Rationality and Self-Interest," in J. Margolis, ed., *The Public Economy of Urban Communities*, Washington 1965.

J. L. Barr and O. Davis, "An Elementary Political and Economic Theory of Local Governments," *Southern Econ. J.*, Oct. 1966, *33*, 149–65.

T. C. Bergstrom and R. Goodman, "Private Demands for Public Goods," *Amer. Econ. Rev.*, June 1973, *63*, 280–96.

W. Birdsall, "A Study of the Demand for Public Goods," in R. A. Musgrave, ed., *Essays in Fiscal Federalism*, Washington 1965.

T. Borcherding and R. Deacon, "The Demand for Services of Non-Federal Governments," *Amer. Econ. Rev.*, Dec. 1972, *62*, 891–901.

A. Downs, *An Economic Theory of Democracy*, New York 1957.

J. E. Grizzle, C. F. Starmer, and G. C. Koch, "Analysis of Categorical Data by Linear Models," *Biometrics*, Sept. 1969, *25*, 489–505.

R. D. Luce, "Semiorders and a Theory of Utility Discrimination," *Econometrica*, Apr. 1956, *24*, 178–91.

P. Shapiro, "A Model of Voting and the Incidence of Environmental Policy," Community and Organization Res. Inst., working pap. no. 1, Univ. California, Santa Barbara 1973.

——— and A. Barkume, "Political Choice and Environmental Quality," in L. Phillips and H. Votey, eds., *Economic Analysis of Pressing Social Problems*, Chicago 1974.

California Secretary of State, *Statement of the Vote: General Election*, Sacramento 1970, 1972.

———, *Official Summary of Arguments on the Propositions, General Election*, Sacramento 1970, 1972.

League of Woman Voters of California, *Pros and Cons of the Propositions, November 7, 1972*, San Francisco 1972.

U.S. Bureau of the Census, *General Social and Economic Characteristics*, Vol. 6, Washington 1972.

Technical Change and the Reproducibility of Capital

By CHARLES R. HULTEN*

The conventional approach to the measurement of total factor productivity distinguishes between shifts in an aggregate production function (technical change) and movements along the production function (factor accumulation).[1] The former is conventionally measured as the residual growth rate of output not accounted for by the growth rates of factor input, and the importance of technical change as a source of economic growth is assessed by comparing this residual with the growth rate of real product.[2]

The conventional approach does not allow for the dynamic interaction of technical progress and capital accumulation. Capital is an output of the production process as well as an input, so that an increase in technical efficiency which leads to an increase in output will in general also lead to an increase in capital stock. This in turn results in a further increase in output, part of which is allocated to capital accumulation, and so on. Thus, when technology is changing, part of the observed capital accumulation must be recognized as having been induced by the changing technology. In assessing the importance of technical change as a source of growth, the effects of induced capital accumulation must be allowed for; the conventional "residual" does not take induced capital accumulation into account, and thus tends to misstate the importance of technical change as a source of growth.

L. M. Read and T. K. Rymes (1968, 1971) have attempted to account for the reproducibility of capital in purely technical terms by measuring capital in terms of the units with which it is produced. J. R. Hicks has proposed to account for reproducibility by measuring capital in terms of consumption foregone. A different approach is taken in this paper: reproducibility is allowed for by introducing intertemporal consumption decisions into the conventional total factor productivity model. Capital accumulation is determined endogenously in this framework, and this permits the analysis of the impact of technical progress on the rate of capital accumulation. It is thus the correct framework for assessing the importance of technical progress in the growth process, while the conventional residual is the correct framework for measuring the change in productive efficiency.[3]

---

* Assistant professor, The Johns Hopkins University. I would like to thank Carl Christ, Dale Jorgenson, Louis Maccini, and Jürg Niehans for their many helpful comments on previous drafts; remaining errors are, of course, my own.

[1] The literature on total factor productivity includes Laurits Christensen and Dale Jorgenson (1969, 1970), Edward Denison (1962, 1967, 1969, 1972), Jorgenson and Zvi Griliches (1967, 1972a,b), John W. Kendrick (1961, 1970), and Robert Solow. Critics of the total factor productivity approach include Luigi Pasinetti, J. R. Hicks, L. M. Read, and T. K. Rymes (1968, 1971).

[2] Actual estimates of the importance of technical change vary. For example, Solow finds that most (seven-eighths) of the growth in *U.S.* real product can be attributed to technical progress (for the period 1909–49). Jorgenson and Griliches (1967), on the other hand, find that nearly all the growth in real product can be attributed to factor accumulation (over the period 1945–65). More recently, Christensen and Jorgenson (1969, 1970) assign a relative importance of approximately one-third the growth rate of real product to technical change (for the period 1929–67).

[3] It is therefore incorrect to conclude, as does Rymes, that "the numbers generated by Denison, Kendrick,

In Section I, the effects of reproducibility are more fully analyzed, as is the relationship between the conventional approach and the approach of this paper. In Section II, a one-sector model of optimal growth is used to investigate the long-run equilibrium expansion of output in response to a changing technology. A rate of technical change (termed the "long-run Fisherian" rate) is defined which measures the total change in output as the economy fully adjusts in the long-run equilibrium to each new level of technology. This corresponds to measuring the shift in the technology along the long-run equilibrium path, as opposed to measuring the shift holding the capital-labor ratio constant, or the capital-output ratio constant, or the labor-output ratio constant. This new definition is the appropriate one for the sources of growth problem, since it measures the shifting technology along the growth path. The new definition is then related to the conventional Hicksian and Harrodian definitions of the rate and bias of technical change. In Section III, the assumption of long-run equilibrium is dropped, and a two-sector model of short-run equilibrium growth is analyzed. The motivation for this change of focus is empirical: the economy cannot be assumed to be on the long-run equilibrium path, if such a path even exists. A short-run index of technical change (termed the "short-run Fisherian" rate) is defined which allows for induced accumulation, and in Section IV this index is estimated using data from the Christensen-Jorgenson (1969, 1970) studies. The result is an estimate of technical change which is nearly twice the Christensen-Jorgenson estimate. While the assumptions underlying the estimation are indisputably crude, the result is at least suggestive of the possible margin by which conventional estimates misstate the contribution of technical change to economic growth.

I. The Reproducibility Problem

To focus more clearly on the issues involved, consider a one-sector economy in which the labor force is constant, and the rate of investment is just sufficient to meet replacement requirements on the existing stock of capital; with constant preferences and technology, the rate of output is stationary. Consider, now, the effect of a once-and-for-all Hicks-neutral shift in technology. The increase in productive efficiency will result in an increase in the rate of output, and the extra output will in general go to increase both consumption and investment. Total investment now exceeds replacement requirements, the capital stock increases, and the rate of output is again increased. The extra output is again divided between consumption and investment, and if the additional investment exceeds the extra replacement requirements generated by the initial increase in investment, the capital stock is further increased and the process is repeated. Each new increment of net investment generates a stream of replacement requirements, and the economy will come to rest (i.e., reach the new equilibrium) when the last increment to gross investment just equals the cumulative additional replacement. In the new equilibrium (b in Figure 1), the rate of output exceeds its previous level (a in Figure 1). The full effect of the change in technology is the amount of output above the initial equilibrium value.

The change in the Hicksian efficiency index from A_0 to A_1 thus results in a total

---

Salter, Solow, Griliches and Jorgenson are meaningless" (1971, p. 2). The total factor productivity framework does in fact retrieve the Hicksian efficiency parameter under the maintained hypothesis of competitive equilibrium and the existence of a Hicks-neutral aggregate production function. The issue confronted in this paper is whether these numbers accurately portray the true importance of technical change in the process of economic growth.

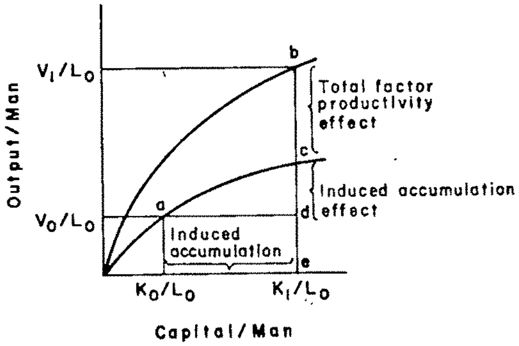


FIGURE 1

increase in output equal to the segment *bd*. The segment *bd* is itself composed of two effects: the change in output due to the increase in factor efficiency (the segment *bc*), and the change in output due to the extra capital induced by the increase in factor efficiency (the segment *cd*). The conventional total factor productivity index measures the ratio *be/ce*, and is thus a valid measure of the ratio of Hicksian efficiencies A_1/A_0. It is not a valid measure of the total importance of technical change, which is measured by the larger ratio *be/de*, because the expansion of capital per man induced by the change in technical efficiency is not allowed for. The magnitude of this induced accumulation effect depends on consumers' preferences for present versus future consumption. The greater the marginal propensity to save, the larger is the induced accumulation effect, and the greater is the margin by which the total factor productivity estimate understates the true contribution of technical change to total economic growth.

II. Long-Run Fisherian Technical Change

These conclusions can be generalized by considering a somewhat more formal model of optimal growth.[4] Consider a one-sector model where preferences for the economy as a whole can be represented by a utility function of the following form:

$$U = \sum_{t=0}^{\infty} a^t u(c_t) \tag{1}$$

where c_t represents per capita consumption, and a is a constant subjective discount factor, $0<a<1$. The function u is assumed to have the following properties

$$u' > 0, \quad u'' < 0 \qquad \text{for } c_t > 0$$

where u' and u'' are the first- and second-order derivatives of u, and to exclude corner solutions

$$\lim_{c_t \to 0} u'(c_t) = \infty$$

K_0 units of the commodity and L_0 units of labor services are initially available, and it is assumed that labor grows exogenously by a constant factor $n \geq 1$: $L_{t+1} = nL_t$. A constant returns technology is also assumed which relates the endowment available at the beginning of each period K_{t+1}, to that part of the previous period's endowment not consumed $K_t - C_t$, and to labor input L_t:

$$K_{t+1} = F(K_t - C_t, L_t) \tag{2}$$

Using lower case letters to denote per capita rates, we may write

$$k_{t+1} = f(k_t - c_t) = \frac{1}{n} F(k_t - c_t, 1) \tag{3}$$

The function f is assumed to have the usual neoclassical properties

$$f' \geqq \frac{1}{n} \qquad f'' < 0 \tag{4}$$

$$f(0) = 0 \qquad f'(0) = \infty \qquad f'(\infty) = \frac{1}{n}$$

where f' and f'' denote the first- and second-order derivatives of f.

The optimal consumption program is the solution to the problem

---

[4] This section follows the exposition of Nissan Liviatan rather closely.

$$\max U = \sum_{t=0}^{\infty} a^t u(c_t)$$

with respect to $c_0, c_1, \ldots$, subject to

$$k_{t+1} = f(k_t - c_t), \qquad 0 \le c_t \le k_t, \qquad t = 0, 1, \ldots, \text{ given } k_0$$

A necessary condition for an optimum is that the marginal rates of substitution in production and consumption must be equal for every t:

$$\text{(5)} \qquad \frac{u'(c_t)}{au'(c_{t+1})} = f'(k_t - c_t) \quad t = 0, 1, \ldots$$

Given assumptions about the discount factor and the concavity of the utility and production functions, condition (5) is also sufficient, provided $a^t u'(c_t) \to 0$ as $t \to \infty$.[5]

The assumptions also insure that the economy can achieve a unique stationary optimum. From (3) and (5), it follows that the stationary value $\bar{k}$ and $\bar{c}$ must satisfy

$$\text{(6)} \qquad \frac{1}{a} = f'(\bar{k} - \bar{c})$$

$$\text{(7)} \qquad \bar{k} = f(\bar{k} - \bar{c})$$

Conditions (6) and (7) completely characterize the stationary optimum. Recalling the example of the preceding section, the equilibria a and b of Figure 1 correspond to solutions of (6) and (7) for two technologies satisfying (4), and for the special case $n = 1$.

We have thus far not allowed for the possibility of technical change. As we are interested in the response of output to changes in technical efficiency, the specification of technology (2) is replaced with one which allows for shifts over time:

$$\text{(2')} \qquad K_{t+1} = F^i(K_t - C_t, L_t) \qquad i = 0, 1, 2, \ldots$$

We assume that each technology F^i, $i = 0, 1, 2, \ldots$, satisfies the conditions (4). Then, for each F^i there is a stationary optimum

$$[\bar{k}_i, \bar{c}_i] \qquad i = 0, 1, 2, \ldots$$

which are the solutions of (6) for each level of technology. The shift in technology from F^i to F^{i+1} has the effect of changing long-run equilibrium output from $\bar{k}_i$ to $\bar{k}_{i+1}$. Given the nature of the utility function (1), this change in output is solely the result of the changing technology, and is therefore equivalent to the Fisherian concept of technical change advanced in this paper. In terms of Figure 1, the ratio $\bar{k}_1/\bar{k}_0$ corresponds to the ratio be/de.

In the case where this technology shifts with sufficient smoothness over time, we can define the path of long-run optimal output $\bar{k}(t)$, which results for the changing technology.[6] A movement along this path indicates technical change in the Fisherian sense of this paper, and may be contrasted with the Hicksian definition of technical change, which is the change in output measured along a constant capital-labor path, and the Harrodian definition, which is the change in output measure along a constant capital-output path.

The rate of change of output along $\bar{k}(t)$ defines the long-run Fisherian *rate* of technical change. The problem is then to characterize this rate, $D\bar{k}(t)/\bar{k}(t)$, in terms of the total factor productivity effect and the induced accumulation effect. This can be done in a straightforward way by including time t as an explicit argument of the technology, and differentiating the second equation of (6) totally with respect to time (see fn. 6). This yields

$$\text{(8)} \qquad \frac{D\bar{k}}{\bar{k}} = \pi \frac{D\bar{\imath}}{\bar{\imath}} + T$$

where $\bar{\imath} = \bar{k} - \bar{c}$, π is capital's share of income, and T is the Hicksian rate of technical change[7] (the operator D denotes derivatives with respect to time, and time subscripts are suppressed for clarity). The left-hand side of (8) is the long-run Fisherian rate of technical change. The second term of the right-hand side T is the conventional total factor productivity effect, and the first term is the effect of induced accumulation.

---

[5] See Liviatan, pages 304–05, for further discussion.

[6] That is, we assume a continuous version of the model in which the technology is characterized by $K(t) = F(I(t), L(t), t)$, where $I(t) = K(t) - C(t)$. The restrictions (4) are assumed to hold for each t, and the optimal path $\bar{k}(t)$ is assumed to be once differentiable.

The marginal product of i (the first equation of (6)) is constant, which implies[8]

$$\frac{Di}{i} = \frac{\sigma}{1-\pi} T + \sigma R \tag{9}$$

where R is the Hicksian bias, and σ is the elasticity of substitution of capital for labor.[9] Substituting (9) into (8), and letting Z denote the long-run Fisherian rate of technical change, we have

$$Z = \left[1 + \frac{\pi}{1-\pi}\sigma\right] T + \sigma\pi R \tag{10}$$

It is immediately evident that in general, $Z \neq T$, implying a divergence between the actual changes in output resulting from a shifting technology Z, and the rate at which the technology is shifting in the sense of Hicks T.

The special case of Hicks-neutral technical change is of particular importance for the conventional analysis of productivity change. The total factor productivity indexes are Divisia line integrals[10] and therefore depend in general on the path of integration. Under Hicks neutrality, however, the line integrals are path independent, and this is consequently the case usually assumed. The implication of Hicks neutrality for the present analysis is that $Z > T$, since $R = 0$ by hypothesis (and $\pi > 0$ and $\sigma > 0$ in the underlying technology). In other words, when the total factor productivity index is independent of the path of integration, it understates the actual change in long-run optimal output, and hence the true contribution of technical change to economic growth.

The long-run Fisherian rate may also be expressed in terms of the Harrodian parameters of technical change. The Harrodian rate H and bias B may be defined in terms of the corresponding Hicksian parameters:[11]

$$\begin{aligned} H &= \frac{T}{1-\pi} \\ B &= [R(1-\pi) + T]\sigma - T \end{aligned} \tag{11}$$

Substitution of H and B for T and R from (11) into (10) yields

$$Z = H + \frac{\pi}{1-\pi} B \tag{12}$$

Therefore, under Harrod-neutral technical change ($B = 0$) the long-run Fisherian rate is equivalent to the Harrodian rate and equals the rate of change of output along the long-run balanced growth path.[12]

---

[7] The share of capital and the Hicksian rate of technical change are defined, respectively, as

$$\pi = \frac{F_K K}{F}, \qquad T = \frac{F_t}{F}$$

where F_K is the partial derivative of F with respect to K, and F_t the partial derivative with respect to t (in the technology of fn. 6).

[8] See Peter Diamond, p. 162, equations (7) and (8), for a description of the dynamic behavior of the marginal products of capital and labor in a one-sector model.

[9] The Hicksian bias and elasticity of substitution are defined, respectively, as

$$R = \frac{F_{Kt}}{F_K} - \frac{F_{Lt}}{F_L}, \qquad \sigma = \frac{F_K F_L}{F F_{KL}}$$

where F_L is the partial derivative of F with respect to L, and F_{Kt}, F_{Lt}, and F_{KL} denote the respective second-order partial derivatives of F with respect to K, L, and t.

[10] For a discussion of Divisia line integrals, see Marcel Richter.

[11] See John Fei and Gustav Ranis.

[12] An alternative approach is to define technical change with respect to the consumption path $\bar{c}$(t), rather than with respect to $\bar{k}$(t), as in this section. With suitable manipulation of (8) and (9), this alternative consumption rate can be characterized in terms of the Hicksian T and R or the Harrodian H and B.

III. Short-Run Fisherian Technical Change

From an empirical standpoint, it is somewhat implausible to assume that the economy moves along the long-run growth path, even if such a path actually exists. In this section, the somewhat weaker assumption is made that the economy is able to achieve a short-run competitive equilibrium at each point in time. In this context, a short-run Fisherian rate of technical change is defined which is the difference between the actual growth rate of real product DV/V and the growth rate of real product which would have occurred if technology had remained constant at the level prevailing in a given base year, DV^*/V^*:

$$(13) \qquad Z' \equiv \frac{DV}{V} - \frac{DV^*}{V^*}$$

Z' measures the growth rate of real product attributable purely to technical change, since DV/V measures the actual growth rate of real product while DV^*/V^* measures the growth rate attributable to the changing factor endowments, excluding the capital accumulation induced by the shifting technology (for this reason Z' is in general different from the Hicksian T). The term Z' is related to the long-run Z in the following way: if technical progress stops after some point in time t″, and if a long-run growth path exists and the economy is on this path at time t′, then the level index associated with Z' (i.e., the integral of the growth rate Z' over the interval [t′, t; t>t″]) converges to the level index associated with Z (the integral of Z over the interval [t′, t; t>t″]). In other words, the short-run and long-run indices of technical change converge.

For purposes of estimation, a two-sector model will be used. This avoids the necessity, implicit in the one-sector case, of imposing a unitary marginal rate of substitution between consumption and investment goods. Separate short-run Fisherian rates are defined for each sector:

$$(14) \quad Z_1' \equiv \frac{DC}{C} - \frac{DC^*}{C^*}, \quad Z_2' \equiv \frac{DI}{I} - \frac{DI^*}{I^*}$$

where C and I are consumption and investment, respectively, and asterisks (here and throughout) denote the values obtaining under a constant base year technology. The aggregate short-run rate of technical change is defined in (24) below, as the weighted sum of Z_1' and Z_2'. Each sector is assumed to possess a constant returns technology,

$$(15) \quad C = G(K_1, L_1, t), \quad I = H(K_2, L_2, t)$$

where $K = K_1 + K_2$ and $L = L_1 + L_2$ denote capital and labor, and t is introduced to allow for shifts in the technologies. Assuming that the economy maintains short-run equilibrium, total differentiation of (15) yields

$$(16) \quad \begin{aligned} \frac{DC}{C} &= \pi_1 \frac{Dk_1}{k_1} + \frac{DL_1}{L_1} + T_1 \\ \frac{DI}{I} &= \pi_2 \frac{Dk_2}{k_2} + \frac{DL_2}{L_2} + T_2 \end{aligned}$$

where $k_1 = K_1/L_1$ and $k_2 = K_2/L_2$, and where π_i and T_i are, respectively, the income share of capital and the Hicksian rate of technical change in the ith sector. Equations (16) constitute the basic model of the total factor productivity approach. All variables in (16) are observable, directly or indirectly, except T_1 and T_2, which are determined residually. The hypothetical C^* and I^* are determined by a similar set of equations, but without technical change

$$(17) \quad \begin{aligned} \frac{DC^*}{C^*} &= \pi_1^* \frac{Dk_1^*}{k_1^*} + \frac{DL_1^*}{L_1^*} \\ \frac{DI^*}{I^*} &= \pi_2^* \frac{Dk_2^*}{k_2^*} + \frac{DL_2^*}{L_2^*} \end{aligned}$$

The problem is to estimate C^* and I^*

given the path of total labor input DL/L, and the initial level of the capital stock $K^0=K_1^0+K_2^0$. We observe first of all that

$$(18)\qquad \frac{DL}{L}=\rho\frac{DL_1^*}{L_1^*}+(1-\rho)\frac{DL_2^*}{L_2^*}$$
$$\frac{DK^*}{K^*}=\theta\frac{DK_1^*}{K_1^*}+(1-\theta)\frac{DK_2^*}{K_2^*}$$

where $\rho=L_1^*/L$ and $\theta=K_1^*/K^*$. Also, assuming that replacement is a constant proportion δ of the capital stock, the basic accumulation equation is

$$(19)\qquad DK^*=I^*-\delta K^*$$

Preservation of equilibrium in the factor markets (i.e., equality of marginal products and factor price ratios) requires, in the absence of technical change, that

$$(20)\qquad \frac{1}{\sigma_1^*}\frac{Dk_1^*}{k_1^*}=\frac{1}{\sigma_2^*}\frac{Dk_2^*}{k_2^*}$$

where σ_i^* is the elasticity of substitution of capital for labor. Finally, the gross savings rate s^* is defined

$$(21)\qquad s^*=\frac{p^*I^*}{C^*+p^*I^*}$$

where p^* is the price of the investment good in terms of consumption. It can be shown that (see Peter Diamond, pp. 162–63):

$$(22)\qquad \frac{Ds^*}{s^*}=(1-s^*)\left[\frac{\pi_1^*(1-\sigma_1^*)-\pi_2^*(1-\sigma_2^*)}{\sigma_2^*}\cdot\frac{Dk_2^*}{L_2^*}+\frac{DL_2^*}{L_2^*}-\frac{DL_1^*}{L_1^*}\right]$$

in the absence of technical change.

If suitable restrictions are imposed on the technology and the gross savings rate s, estimates of the factor growth rates in each sector may be derived from (18)–(22). These estimates can then be used in (17) to calculate the growth rates of C^* and I^*. The simplest assumptions with which to work are

$$(23)\qquad s=\text{constant},\qquad \sigma_1=\sigma_2=1$$

This, of course, implies that the factor shares π_i are constant. Although these assumptions are arbitrary, they are not without empirical appeal. Alternatively, a more flexible technological specification, such as the *CES*, could be used, in which case

$$\frac{D\pi_i^*}{\pi_i^*}=(1-\pi_i^*)\left(1-\frac{1}{\sigma_i^*}\right)\frac{Dk_i^*}{k_i^*}\qquad i=1,2$$

For purposes of estimation, the assumptions (23) will be employed. It then follows immediately that

$$\frac{DL}{L}=\frac{DL_1^*}{L_1^*}=\frac{DL_2^*}{L_2^*},\quad \frac{DK^*}{K^*}=\frac{DK_1^*}{K_1^*}=\frac{DK_2^*}{K_2^*}$$

The growth rate of labor DL/L is given, and DK^*/K^* can be computed recursively from (19), given the initial value of K. Thus with constant factor shares,

$$\frac{DC^*}{C^*}=\pi_1\frac{Dk^*}{k^*}+\frac{DL}{L}$$

Given these estimates, the calculation of Z_1' and Z_2' is straightforward. An overall rate of technical change may be obtained by weighting the sector rates by the actual sector shares in total product:

$$(24)\qquad Z'=(1-s)Z_1'+sZ_2'$$

Finally, the conventional indexes of technology T_1 and T_2 can be subtracted from Z_1' and Z_2' to estimate the effects of induced capital accumulation.

Thus the conventional dichotomy between technical change and capital accumulation results in the following aggregate relationship:

$$(25)\qquad \frac{DV}{V}=\pi\frac{Dk}{k}+\frac{DL}{L}+T$$

where V is total real product, π the share of capital in total product, and $T=(1-s)T_1+sT_2$. The short-run Fisherian approach, on the other hand, results in a reallocation of the sources of growth:

$$\frac{DV}{V}=\pi\frac{Dk^*}{k^*}+\frac{DL}{L}+Z' \tag{26}$$

which takes into account the effect of the induced accumulation of capital (assuming, as before, constant shares).

IV. Empirical Application

The model was estimated using the Christensen and Jorgenson (1969, 1970) data for the private domestic economy. The results are presented in Tables 1 and 2. For the period 1948–66, short-run Fisherian technical change was found to have occurred at an average annual rate of 2.67 percent. This is nearly twice the 1.42 percent average annual growth rate of total factor productivity T, implying an induced accumulation effect of 1.25 percent per year. The average annual sector rates Z_1' and Z_2' were found to be 2.25 and 3.22 percent, respectively. Results for 1950–63 are similar. Unfortunately, estimates of the corresponding total factor productivities T_1 and T_2 are not available.

The importance of technical change as a source of economic growth is correspondingly higher in the new view. The Fisherian rate accounts for approximately 64 percent of the growth rate of real product. This compares with 34 percent for the total factor productivity rate. The increased importance assigned to technical change is, of course, due to the treatment of capital input as a source of growth. Over the period 1948–66, capital actually grew at an average annual rate of 4.22 percent. The projected rate of capital accumulation in the absence of technical change is estimated to have been only 1.49 percent per year.[13] Consequently, 2.73 percentage points of the actual growth rate of capital is attributed to induced accumulation, and assigned to technical change.

It is thus evident that the induced accumulation effect is a potentially important source of economic growth. Under the present assumptions, the effect attributable to induced accumulation is nearly equal in magnitude to the conventional total productivity effect. This conclusion does obviously depend on the validity of the assumptions: the constancy of the sav-

TABLE 1—AVERAGE ANNUAL PERCENTAGE GROWTH RATES O- PRODUCTIVITY

Variable	1948–66	1950–66
Total Factor Productivity T	1.42	1.33
Fisherian Factor Productivity, Z'	2.67	2.51
Fisherian Factor Productivity (C), Z_1'	2.25	2.17
Fisherian Factor Productivity (I), Z_2'	3.22	2.93

TABLE 2—SOURCES OF GROWTH OF REAL PRODUCT, 1948–66

(Average Annual Percentage Growth Rates)

	Christensen-Jorgenson	Fisherian
Real Product	4.15	4.15
Real Factor Input	2.73	1.48[a]
Capital	4.22	1.49
Labor	1.50	1.50
Technical Change	1.42	2.67
Relative Proportion of Technical Change	0.34	0.64

[a] The growth rate of real factor input should be a weighted average of the growth rates of capital and labor. The fact that the rate for real factor input is slightly below the input rates is due presumably to round-off error.

---

[13] The capital stock projections assume an average rate of replacement δ equal to 0.083. This is a weighted average of the Christensen and Jorgenson (1969, 1970) rates of replacement for plant and equipment. The value of π_1 was estimated from the share of capital in aggregate income π and the gross savings rate s, and found to be approximately 0.31. Alternative estimates of capital stock were projected using values for π_1 of 0.38 and 0.45, but the resulting differences were negligible.

ings rate and unitary elasticities of substitution. However, these assumptions are certainly not implausible, or at least not so implausible as to completely vitiate the empirical analysis. This would imply that a reassessment of the role of technical change in the growth process is in order.

Different rates of total factor productivity are found in other investigations. For the period 1950–62, Edward Denison (1969) obtains an average annual rate of 1.37 percent, Jorgenson and Griliches (1967) a rate of 0.30 percent, and Kendrick (1970) a rate of 2.01 percent, for the private domestic economy. For the same period, Christensen and Jorgenson find a rate of 1.33 percent per year, and the current estimate, gross of induced accumulation, is 2.51 percent per year. However, this latter estimate is not directly comparable to the alternate estimates of total factor productivity, since it is specific to the Christensen-Jorgenson data base, and would have to be altered in order to reflect the methodologies of the other studies. There is no reason to believe that the broad conclusions of this paper would be changed if this were done.

V. Conclusion

The problem of measuring disembodied technical change includes both the measurement of total factor productivity and the measurement of the importance of technical change as a source of economic growth. Existing studies have successfully dealt with the first aspect of the problem, but not the second. The importance of technical change depends not only on the change in total factor productivity, but also on the response of capital (and perhaps labor also)[14] to the change. When these induced effects are allowed for, technical change is seen to be potentially far more important than heretofore thought. While the estimates of this paper are based on somewhat arbitrary assumptions, they are at least suggestive of the possible magnitude by which the conventional total factor productivity approach understates the importance of technical progress.

REFERENCES

L. R. Christensen and D. W. Jorgenson, "The Measurement of U.S. Real Capital Input, 1929–1967," *Rev. Income and Wealth,* Dec. 1969, *15,* 293–320.

——— and ———, "U.S. Real Product and Real Factor Input, 1929–1967," *Rev. Income and Wealth,* Mar. 1970, *16,* 19–50.

E. F. Denison, *The Sources of Economic Growth in the United States and the Alternatives Before Us,* New York 1962.

———, *Why Growth Rates Differ: Postwar Experience in Nine Western Countries,* Washington 1967.

———, "Some Major Issues in Productivity Analysis: An Examination of the Estimates of Jorgenson and Griliches," *Surv. Curr. Bus.,* May 1969, *49,* 1–28.

———, "Final Comments," *Surv. Curr. Bus.,* May 1972, *52,* 95–110.

P. A. Diamond, "Disembodied Technical Change in a Two-Sector Model," *Rev. Econ. Stud.,* Apr. 1965, *32,* 161–68.

J. C. H. Fei and G. Ranis, "Innovational Intensity and Factor Bias in the Theory of Growth," *Int. Econ. Rev.,* May 1965, *6,* 182–98.

J. R. Hicks, *Capital and Growth,* London 1965.

D. W. Jorgenson and Z. Griliches, "The Explanation of Productivity Change," *Rev. Econ. Stud.,* July 1967, *34,* 249–83.

——— and ———, (1972a) "Issues in Growth Accounting: A Reply to Edward F. Denison," *Surv. Curr. Bus.,* May 1972, *52,* 65–94.

——— and ———, (1972b) "Final Reply," *Surv. Curr. Bus.,* May 1972, *52,* 111.

---

[14] It is also possible that population growth and labor force participation rates are influenced by technical change. This is particularly plausible for population growth, if sufficiently long time periods are considered.

J. W. Kendrick, *Postwar Productivity Trends in the United States, 1948–1969,* New York 1970.

———, *Productivity Trends in the United States,* Princeton 1961.

N. Liviatan, "A Diagrammatic Exposition of Optimal Growth," *Amer. Econ. Rev.,* June 1970, *60,* 302–09.

L. L. Pasinetti, "On Concepts and Measures of Changes in Productivity," *Rev. Econ. Statist.,* Aug. 1959, *1,* 270–86.

L. M. Read, "The Measurement of Total Factor Productivity Appropriate to Wage-Price Guidelines," *Can. J. Econ.,* May 1968, *1,* 349–58.

M. K. Richter, "Invariance Axioms and Economic Indexes," *Econometrica,* Oct. 1966, *34,* 739–55.

T. K. Rymes, *On Concepts of Capital and Technical Change,* Cambridge 1971.

———, "Professor Read and The Measurement of Total Factor Productivity," *Can. J. Econ.,* May 1968, *1,* 359–67.

R. M. Solow, "Technical Change and The Aggregate Production Function," *Rev. Econ. Statist.,* Aug. 1957, *39,* 312–20.

Cross-Subsidization: Pricing in Public Enterprises

By GERALD R. FAULHABER*

The problem of pricing commodities produced in the presence of common costs by a publicly owned or regulated enterprise continues to be of great practical interest to regulators and policy makers. It has generated a rich literature in the fields of public finance and public utility regulation, dating back to the 1880's. The economics literature has by and large focused on the efficiency of such commodity prices, whereas public policy makers are also concerned about more loosely defined questions of equity: does a proposed price structure for the multicommodity enterprise "unduly" favor the consumers of one commodity at the expense of the purchasers of another commodity, i.e., does the price structure result in cross-subsidy?

A paradigm best serves to illustrate the problem: a profit-regulated railroad which interconnects two large cities also provides rail service to a smaller town located on the route between the two cities. The fares charged for passage from the small town are sufficient to generate revenues in excess of the additional costs of serving it, such as ticketing and station costs, but not sufficient to cover an equal or proportionate (however defined) share of the common costs, such as trackage, signaling, and trainyard costs. Since the small town is not paying its "fair share" of common costs, is it not being subsidized by the larger cities? Do not the railroad tariffs favor the town at the expense of the cities?

Provided the revenues realized from providing rail service to the town exceed the added costs, the answer must be in the negative. In fact, the provision of service under these circumstances (assuming the road's profits remain constant) may actually lead to lower fares for the two cities compared to the alternative of no service for the town at all. The principle was stated by E. Porter Alexander: "If the [New York to San Francisco] price is more than the additional outlay in doing it, as against leaving it alone, it is profitable to the railroad, and the business is moreover advantageous to the whole inland community served by the railroad. For . . . the more prosperous the road, the lower the local rates may be made" (p. 4). The argument is no less cogent today, as William Baumol and David Bradford point out: "For (particularly if the firm is subject to a constraint on its overall profit) the opening of a market which makes any net contribution [to common costs] may permit or may even require a reduction in prices elsewhere" (p. 278).

As a first approximation, we may use this intuitively appealing notion as the basis for a definition: If the provision of any commodity (or group of commodities) by a multicommodity enterprise subject to a profit constraint leads to prices for the other commodities no higher than they would pay by themselves, then the price structure is *subsidy-free*. Thus, a subsidy-free price structure insures that the provision of each commodity by the enterprise is "Pareto superior" to nonprovision.[1] Note, however, that a subsidy-free price structure is not necessarily welfare maximizing; nor are we entitled to assume that such price structures are morally superior to their subsidy-prone fellows on grounds of social justice. Subsidy-free prices do no more than insure that the production and sale of each commodity makes all consumers at least as well off as they would otherwise be.

---

* Bell Telephone Laboratories, Inc. I wish to express my thanks for the helpful comments and suggestions of William Baumol, David Bradford, and Michael Rothschild of Princeton University. I am also indebted to my colleagues at Bell Laboratories, especially Elizabeth Bailey and V. S. Bawa for their very useful criticisms. This work was inspired by the earlier research of Edward Zajac, to whom I am especially indebted for continuing discussions on this work. Of course, responsibility for errors is mine.

We shall see, however, that our subsequent extension of the definition of subsidy-free prices as the solution set to an n-person cooperative game leads to rather strong conclusions concerning the stability of a competitive market, the cogency of the "natural monopoly" argument, and the choice between protected monopoly and open competition.

I. The Enterprise as a Cooperative Game

We restrict our attention to enterprises which (a) produce and sell at least two differentiable products which can be priced separately, and (b) operate under a constraint on total profit requiring total revenues to exceed the sum of the added costs of each commodity. The natural candidates to fulfill conditions similar to (b) are enterprises characterized by economies of joint production[2] subject to a break-even constraint, such as privately owned publicly regulated utilities, or a publicly owned enterprise required to "stand on its own bottom" (to use Harold Hotelling's mixed metaphor), such as a bridge or turnpike authority.[3] Without loss of generality, we may assume that total profits (earnings in excess of the cost of capital) are constrained to be zero.[4,5]

As an example, suppose we consider the supply of water to a residential community, when commodity i is water delivered to neighborhood i (or perhaps household i).[6] We denote the set of neighborhoods (commodities, consumer groups) by $N=\{1, 2, \ldots, n\}$, where $q_1, \ldots, q_n$ are the amounts of water supplied to each neighborhood. We adopt the notation that for all subsets $S\subseteq N$, q^S is an n-vector for which

$$q_i^S = \begin{cases} q_i & i \in S \\ 0 & i \in N - S \end{cases}$$

so that $q^N=(q_1, q_2, \ldots, q_n)$ and $q^{\{1\}}=(q_1, 0, \ldots, 0)$.

There are many production arrangements by which water can be supplied: each neighborhood can have its own well, storage tank, and pumps; groups of neighborhoods can have common facilities; or the entire community can use one common facility. The technology of water recovery, storage, and distribution we consider to be available to all, and representable by a cost function $C(\cdot)$ defined over n-vectors, i.e., the cost of supplying water to the set $S\subseteq N$ of neighborhoods is $C(q^S)$. Thus, if neighborhood 1 constructed its own well, pumps, and storage tank, the cost would be $C(q^{\{1\}})=C(q_1, 0, \ldots, 0)$, and the total cost if all neighborhoods were to supply themselves would be $\sum_{i=1}^{n} C(q^{\{i\}})$. At the other extreme, if a single enterprise served all neighborhoods, the total cost would be $C(q^N)$. Our assumption of economies of joint production is expressed in cost function terms by requiring that $C(\cdot)$ satisfy for all S, $T\subseteq N$, with $S\cap T=\phi$, and any $q_i\geqq 0$

---

[1] In the sense used by James Quirk and Rubin Saposnick (1968): each consumer finds the state in which each commodity is provided by the enterprise to be preferable to the state in which only some of the commodities are provided by the enterprise. This interpretation was pointed out to me by Baumol.

[2] One could imagine a government-owned firm with constant returns to scale, but where the firm is required to contribute a fixed sum annually to the state's coffers.

[3] The imposition of the profit constraint, of course, precludes strict marginal cost pricing. An extensive literature, beginning with Hotelling, suggests the desirability of marginal cost pricing with public subsidies to cover deficits in increasing return industries; the difficulties affecting this approach are well known, and concisely discussed by Nancy Ruggles. As has often been pointed out, it is not easy to devise a tax or subsidy which will not affect the marginal conditions. Clearly, our problem lies in the area of the second best.

[4] We assume throughout that the enterprise operates efficiently, and is regulated or managed to just earn its cost of capital. There is an extensive literature in public utility regulation which addresses the difficulties of achieving what we so blithely assume, treated most recently and most elegantly by Elizabeth Bailey. Such issues are not the focus of this paper.

[5] We may interpret positive profits as a subsidy of producers (investors) by consumers, and negative profits as a subsidy of consumers by producers. Thus, our zero-profit constraint may itself be viewed as a requirement of subsidy-free operation. As previously asserted, however, the analysis requires only that the profit constraint satisfy condition (b) above

[6] Although in our example the output commodities differ only in the location to which they are delivered, the analysis holds in more general circumstances; for example, a dam can supply three commodities: flood control, water for irrigation, and electric power.

$$C(q^S) + C(q^T) \geqq C(q^{S+T}) \tag{1}$$

That is, the cost of supplying water to two groups of neighborhoods jointly is less than separate supply, at any demand levels q_i.

On the demand side, we assume that quantities demanded are a function of own price only,[7] i.e., $q_i=q_i(p_i)$, where $p_i=$ price of the ith commodity. To maintain consistent notation, we denote the revenues from neighborhood i as $r_i=p_i\cdot q_i(p_i)$, and

$$R(q^S) = \sum_{i\in S} r_i = \sum_{i\in S} p_i\cdot q_i(p_i)$$

Once a particular supply arrangement has been determined, then the suppliers are held to a zero-profit constraint. For example, if the set $S\subseteq N$ is served by one supplier who supplies set S only, then the prices p_i, $i\in S$, must satisfy

$$R(q^S) - C(q^S) = \pi(q^S) = 0 \tag{2}$$

where the above equation (2) defines the profit function $\pi(\cdot)$.

In view of inequality (1), it is clear that the arrangement in which one producer supplies the entire community at the levels q^N is at least as efficient as any other arrangement, and that if the inequality is strict, the single supplier is the uniquely most efficient production arrangement.[8] From the point of view of the application of these concepts, we are interested in prices which will induce and preserve the "coalition of the whole," i.e., the single supplier arrangement. Thus, the relevant constraint is

$$R(q^N) - C(q^N) = \pi(q^N) = 0 \tag{2'}$$

The reader who is familiar with the theory of n-person cooperative games (see, for example, Robert Luce and Howard Raiffa) may recognize the structure of our "game"; for *given demand levels* $q_1, \ldots, q_n$, the neighborhoods (or consumer groups) are the "players," the function $C(\cdot)$ is the characteristic function, denoting the value of the game as played by all possible coalitions $S\subseteq N$ of players, and the vector of revenues $\{p_iq_i\}$ is the "payoff" vector.[9] Expression (1) is the condition of subadditivity, which insures that there are incentives to cooperate. The economic interpretation is that of a natural monopoly, a classic case for public control (see, for example, Alfred Kahn). Equation (2), our zero-profit constraint, corresponds to the game theoretic condition that all the "value" of the games be shared amongst the players via the payoff vector.

---

[7] The more interesting and more difficult case of interrelated demand functions is treated later in the paper. Also note that a single price prevails, levied on each unit consumed. We do not treat the case of a multipart tariff, e.g., a connection charge plus a consumption charge.

[8] The zero-profit assumption is most crucial here.

[9] In the standard game theory formulation, payoffs are received by the players, whereas in our game, revenues are paid out by the players. The correspondence may be made formally complete by changing signs (e.g., payoffs are negative revenues) and thus reversing the direction of all inequalities.

For a given subadditive characteristic (cost) function $C(\cdot)$, when is a payoff vector $r_1, \ldots, r_n$ a "solution" to the game? Loosely speaking, a payoff vector is a solution if it induces all (rational) players to cooperate, thus achieving (by expression (1)) the highest valued (least cost) game. At the given demand levels, if each neighborhood i is to be induced to join, then prices p_i must satisfy

$$(3) \qquad R(q^S) \leqq C(q^S) \qquad \forall S \subseteq N$$

That is, in order to induce each neighborhood $i \in N$, and each group of neighborhoods $S \subseteq N$ to cooperate, the community water company's prices should be such as to make it more costly for any group of neighborhoods to "go it alone."[10] In the parlance of game theory, revenue vectors which satisfy (2′) and (3) constitute the core of the game, as defined by D. B. Gillies. We can now formalize our definition of subsidy-free prices: prices for which the resulting revenue vector lies in the core of the game are subsidy-free.

Another interpretation of subsidy-free prices may be obtained by subtracting (3) from (2′) to obtain

$$(3') \qquad R(q^S) \geqq C(q^N) - C(q^{N-S}) \qquad \forall S \subseteq N$$

That is, the revenues contributed by the set of neighborhoods $S \subseteq N$ should be at least as great as the added cost of supplying S, in order to be subsidy-free.

A simplified version of (3′) involving only single-element sets has been known in the public utility field for some time, and has been called the incremental cost test. Its lineage is ancient, dating back to Alexander, as we have seen, and was recently restated by Kahn. However, this test for cross-subsidy is generally used to examine a single product or service, rather than groups of products, which, as we shall see below, can be misleading.

We might also interpret (3) as a "stand-alone" test, and although it is logically equivalent to (3′) when used in conjunction with (2′), it has received almost no attention in the literature of regulation or public utility economics. Exceptions are Karl Borch in his game theoretic treatment of automobile insurance rates in which expressions (2′) and (3) play a role, and a reference by Harry Trebing to rate maxima determined by such a stand-alone criterion.[11]

An example will serve to illustrate these ideas, in particular the importance of examining all subsets $S \subseteq N$ rather than only individual consumer groups $i \in N$. Suppose our water company serves four neighborhoods, with neighborhoods 1 and 2 east of the well, and neighborhoods 3 and 4 west of the well. The cost of the well and storage tank is \$200; in addition, the firm has an eastbound trunk line and pumping station for q_1 and q_2, as well as similar facilities devoted to q_3 and q_4, each of which costs \$100. In addition, the distribution costs for delivery of $q_i = 10{,}000$ gallons/day to each neighborhood are \$100, comprised of \$80 for the local distribution mains, plus \$.002 pumping costs, or \$20. Thus, if all neighborhoods are served by the single water company, total costs are \$800, or

---

[10] Of course, if neighborhood i actually did go it alone, the supply arrangement would change and prices would have to change, since (3) would not satisfy the zero-profit constraint. In our example with fixed (inelastic) demands, if $r_i \leqq C(q^{\{i\}})$ and yet neighborhood i tried to go it alone, its price would rise from p_i to p_i', where

$$p_i' - p_i = \frac{C(q^{\{i\}}) - r_i}{q_i} \geqq 0$$

[11] Trebing states: "A more meaningful standard might relate maximum rates to the cost of a single purpose facility or system built to serve the user requirements of the particular group most affected by the upper price limit. If this group paid a rate in excess of the cost of the single purpose facility, it would be subsidizing other user groups as well as failing to participate in any of the economies of joint-cost inherent in public utility operation" (p. 69).

$$C(q_1, q_2, q_3, q_4) = \$800$$

If any one neighborhood i were to go it alone, the only costs incurred would be the well, tank, and distribution system, so that

$$C(q_1, 0, 0, 0) = C(0, q_2, 0, 0) = C(0, 0, q_3, 0) = C(0, 0, 0, q_4) = \$300$$

Any two adjacent neighborhoods, i.e., 1 and 2, or 3 and 4, need only construct the well, tank, and two distribution systems, so that

$$C(q_1, q_2, 0, 0) = C(0, 0, q_3, q_4) = \$400$$

Two (three) nonadjacent neighborhoods would find it necessary to build well, tank, two (three) distribution systems, and a trunk line and pumping station to the nonadjacent neighborhood, yielding

$$C(q_1, 0, q_3, 0) = C(q_1, 0, 0, q_4) = C(0, q_2, q_3, 0) = C(0, q_2, 0, q_4) = \$500$$

$$C(q_1, q_2, q_3, 0) = C(q_1, q_2, 0, q_4) = C(q_1, 0, q_3, q_4) = C(0, q_2, q_3, q_4) = \$600$$

What revenue vectors are subsidy-free? Using (2′), (3), and (3′), some manipulation will show that the relevant relations are

$$\begin{aligned} r_1 + r_2 + r_3 + r_4 &= \$700 \\ r_i &\geqq 100 \\ r_1 + r_2 &\geqq 300 \\ r_3 + r_4 &\geqq 300 \end{aligned}$$

Will the time-hallowed incremental cost test applied singly to each service ($r_i \geqq 100$) be sufficient to detect subsidy? Unfortunately, no; suppose the price per gallon to the eastern neighborhoods is 1.2¢, and to the western neighborhoods is 2.3¢. Then $r_1 = r_2 = \$120$, and $r_3 = r_4 = \$230$. Clearly, total revenues just equal total costs, and each neighborhood is paying more than its incremental cost of \$100. However, it is clear that neighborhoods 1 and 2 together, the eastern neighborhoods, are not paying in total their joint incremental cost of \$300, which includes not only the two distribution systems, but the eastbound trunk line and pumping station costs of \$100. Their combined revenue of \$240 falls short of their combined incremental cost, i.e., $r_1 + r_2 < 300$. The problem arises because the trunk line and pumping costs are joint among a subset of neighborhoods, larger than a single neighborhood and smaller than the entire collection of neighborhoods.[12] If all costs were either directly attributable to a single neighborhood (i.e., could be avoided were that neighborhood to drop out of the coalition) or were joint among all neighborhoods, then an incremental cost test for each service taken one at a time would suffice. The presence of costs which are joint among a proper subset of the whole coalition imply the necessity of testing that subset for cross-subsidy.[13]

---

[12] A similar calculation shows the same difficulties in applying the "stand-alone" test of (3). If each neighborhood tried to go it alone, they would individually incur well, tank, and distribution costs of \$300 each, which is greater than the revenues of any of the neighborhoods under price structure (9), thus satisfying (3). However, if the western neighborhoods decided to "go it alone, together," their total stand-alone costs would be \$400, which is less than their combined \$460 expenditure.

[13] The application of game theory to public enterprise pricing has been the subject of several recent articles. S. C. Littlechild, Edna Loehman and Andrew Whinston (1971, 1974a,b) recommend the use of the Shapley value of the appropriately defined game, on the basis of its purported "fairness" and "equity." In this paper, the emphasis is not on finding a unique set of prices which is "fair," but rather on determining a set of prices, all of which are subsidy-free, and, as we discuss below, provide the appropriate incentives for consumer groups to seek the most efficient means of supply in the presence of joint production. This leads us to the core as a solution concept; as the Shapley value need not lie in the core of the game, we find it of less interest than the cited authors. Our approach is much closer in spirit to the literature in club theory, such as James Buchanan and Mark Pauly (1967, 1970). Pauly considers core solutions as club equilibria, i.e., as a means of inducing and preserving optimal size clubs. In our case, core solutions have similar incentive properties, and the

Although prices are our major interest, the analysis so far has treated them as simply a way to obtain revenues r_i from fixed quantities q_i, where the revenues play the key role in the core constraints. Suppose that both (2′) and (3) were satisfied for some demand levels $q_1, \dots, q_n$, at prices $p_1, \dots, p_n$, but that some set $S \subseteq N$ of consumers thought they would obtain lower prices by going it alone. This would change the supply structure so that one enterprise supplied the rebellious consumers of S, while presumably another enterprise supplied $N-S$. The zero-profit constraint would apply to both so that in view of (3), prices and hence quantities need adjust accordingly to $\bar{p}_1, \dots, \bar{p}_n$ and $\bar{q}_1, \dots, \bar{q}_n$. At the new demand levels it may be the case that $\bar{p}_i \leqq p_i$ for $i \in S$, even though (3) holds at the old demand levels. Thus, consumers in S may be better off by going it alone even though $r_1, \dots, r_n$ is in the core.

We show that this cannot occur, provided that the function $\pi(q^S) = R(q^S) - C(q^S)$ has a positive derivative with respect to price p_i for all $i \in S$.[14] To prove this, we will show that uniformly lower prices for any set $S \subseteq N$ cannot satisfy the zero-profit constraint for that set S provided (3) is satisfied and the derivative of $\pi(q^S)$ with respect to p_i for all $i \in S$ is positive.

Suppose that at prices $p_1, \dots, p_n$, (2′) and (3) are satisfied for all $S \subseteq N$, and the derivatives of $\pi(\cdot)$ are positive, as assumed. For any set $S \subseteq N$, let $\bar{p}^S$ be a price vector for S such that for all $i \in S$ $\bar{p}_i \leqq p_i$, and for at least one $j \in S$, $\bar{p}_j < p_j$. Since (3) is satisfied at prices $p_1, \dots, p_n$ and quantities $q_1, \dots, q_n$ and all cross-elasticities are zero, then

$$
\begin{aligned}
0 \geqq \pi(q^S) &= \pi\{q(\bar{p}^S)\} + \sum_{i \in S} \int_{\bar{p}_i}^{p_i} \frac{\partial \pi}{\partial x_i}\, dx_i \\
&> \pi\{q(\bar{p}^S)\} = R(\bar{q}^S) - C(\bar{q}^S)
\end{aligned}
$$

Hence, for any $S \subseteq N$ and any set of prices $\bar{p}^S$ which would induce S to go it alone, the resulting coalition would be infeasible in that revenues would not cover costs for the rebellious coalition S.

Clearly, this result is quite helpful as a practical matter to policy makers concerned with determining whether or not a given rate structure is subsidy-free. When considering if any consumer group would be better off on a stand-alone basis, the regulator need not calculate a new set of prices and demands which would obtain under the hypothetical stand-alone arrangement, but need only calculate revenues based on the given price structure and demand levels, and the costs of the hypothetical arrangement based on the given demand levels (assuming, of course, the conditions of the theorem are fulfilled). One needs no information about demand elasticities but rather only information about the costs of alternative supply arrangements at the given demand levels to

---

optimal size "club" is the entire set of players, since the economies of joint production are assumed to prevail up to the entire set N. My work differs from Pauly's principally in that I am concerned with pricing rules for divisible commodities rather than entry fees to a club. Further, Pauly's major results are for the case of identical consumers, although he generalizes this in his 1970 paper, where the optimal club size is small relative to the size of the population as a whole. In the present paper, not only are consumers different on the demand side, but also on the supply side; the cost of serving different consumer groups may differ. In addition, I concentrate on the case where economies of joint production encompass the entire relevant population.

[14] We argue for the plausibility of this assumption as follows: suppose $\pi(q^s)$ as a function of p_i is concave and achieves a maximum at some positive level of profits. If the profit constraint is binding, then $\partial\pi(q^S)/\partial p_i \neq 0$. The firm's constrained maximum could occur at a lower price, where $\partial\pi(q^S)/\partial p_i > 0$, or at a higher price, where $\partial\pi(q^S)/\partial p_i < 0$ (by concavity of π), and a profit-maximizing firm would in theory be indifferent between the two points. However, at the lower price, output would be higher, sales (it turns out) would be higher, and consumers happier. Since all are at least as well off at the lower price, the firm would have to be perverse indeed to operate where $\partial\pi(q^S)/\partial p_i < 0$. Hence, we may safely assume, barring ignorance or malice, that $\partial\pi/\partial p_i > 0$.

determine whether or not prices are subsidy-free in the zero cross-elasticity case.

II. Market Incentives and Welfare Maximization: Some Implications

Before moving on to the technically more difficult discussion of cross-subsidization when demands for the commodities are interrelated, we explore first the issue of incentives to competitive entry and subsidy-free prices; second the relation of welfare maximizing prices to subsidy-free prices; and third the question of existence of subsidy-free prices.

Recall that our definition of subsidy-free prices as the core of the game focusses on the inducement aspect: a price structure which is subsidy-free is Pareto superior to what prices would otherwise be for each consumer group or coalition of consumer groups. Thus, all groups can in principle be induced to join the overall (and most efficient) coalition, since the benefits of increasing returns are distributed in some measure to each group. A public enterprise with economies of joint production extending to the entire market and whose prices lie in the core is the classic case of a natural monopoly;[15] the distribution of the economies of joint-cost operation are such that no firm would find it profitable to enter the market for some subset of the enterprise's consumers in view of (3). If there are economies of joint operation and prices are not in core, however, then at least one coalition will have strong incentives to find themselves an entrepreneur and split off from the enterprise, thus leading to a globally less efficient supply arrangement. If free entry into the market is the rule, then if the public enterprise chooses prices outside the core, market forces should be expected to lead to the virtual breakup of the enterprise!

As pointed out before, prices which are subsidy-free do not necessarily promote the common weal or bring about social justice. The regulators of our public enterprise may deliberately choose a policy of subsidization, and price the commodities of that enterprise accordingly. Our analysis indicates that such prices would produce market incentives for the erosion of the firm offering the economies of joint production by competitive entry. If the public authorities desire both subsidizing prices *and* economies of scale, then the coercive authority of the government must be employed to restrict or prohibit entry into the market. Thus, even when the public enterprise enjoys increasing returns to scale, if the regulators adopt a pricing policy of subsidization as we have defined, entry must be restricted.[16] On the

---

[15] Typically, natural monopoly is discussed in terms of an enterprise producing only one commodity with increasing returns to scale. This situation is simply a special case of (1), since with a single commodity, for any arbitrary partition of the consumers into n groups, we can define our cost function

$$C(q_1, \ldots, q_n) = C\left(\sum_{i=1}^{n} q_i\right) = C(q)$$

Condition (1) then states that it is less costly to have one enterprise produce the entire supply of the single commodity than to split the supply function among two or more producers, which is precisely the meaning of natural monopoly.

[16] At least two caveats must be added to this discussion of competitive entry, both relating to the limitations of this static analysis when applied to actual pricing decisions. First, it may be the case that the opportunities for alternative supply arrangements may change drastically once a decision has been reached on a particular cooperative arrangement. For example, prior to the construction of a dam, those who dwell in the river's flood plain have the option to provide flood control through a distributed network of upstream catch-pools. Once the dam is built, however, the lake which builds up behind the dam may preclude the catch-pool method of flood control. Consumer groups whose options for alternative supply are few *ex post* can perhaps be expected to seek long-term contracts or commitments *ex ante*, when they have more options for alternative supply, and hence a more favorable bargaining position.

The second problem arises when technological innovation changes the nature of the game. Dynamic changes of this nature may affect not only the level of costs, but also the cost structure of the industry, turn-other hand, if the price structure is subsidy-free, then no such protection is needed since competitive entry is not feasible in view of (3).

Suppose that a regulator deliberately adopts a policy of cross-subsidization in his pricing decisions. Would such regulation necessarily violate some principle of welfare maximization, thus leading to subsidization? Indeed not; as it shall be shown, there is no a priori reason to expect that prices which maximize welfare subject to a break-even constraint will necessarily be subsidy-free. Intuitively, we might expect this since differences among constrained welfare-maximizing prices depend on differences in demand elasticities as well as costs, whereas whether prices are subsidy-free or not depends on the costs of alternative supply arrangements.

To illustrate this point, let us return to our water company, where the managers/regulators are maximizing welfare subject to a profit constraint, à la Frank Ramsey, M. Boiteux, and Baumol-Bradford. Specifically, we assume our water company: (a) views the incomes of consumers and the prices of all other commodities as fixed; (b) is required to break even; (c) faces zero cross-elasticities of demand among the different neighborhoods; and (d) wishes to maximize welfare subject to (a), (b), and (c). It is well known that prices thus determined have their percentage deviations from marginal cost inversely proportional to the price elasticity of demand, which prices Baumol and Bradford call "quasi optimal." That is,

$$(4) \qquad \frac{p_i - MC_i}{p_i} = \frac{k}{\epsilon_i} \qquad i = 1, 2, \ldots, n$$

where ϵ_i is the elasticity of demand with respect to price for the ith consumer group, and MC_i is the marginal cost of commodity i.

Returning to our previous example of the community water company, the relevant marginal cost is the pumping cost per gallon of \$.002. Further, assume that the consumers of water in neighborhoods 3 and 4 are more insistent upon watering their lawns than the consumers in neighborhoods 1 and 2, so that[17] $\epsilon_1 = \epsilon_2$ and $\epsilon_3 = \epsilon_4$, but that $\epsilon_3/\epsilon_1 = \epsilon_4/\epsilon_2 = .91$, i.e., groups 3 and 4 have a less elastic demand than groups 1 and 2. With this not very great difference in elasticities, and a marginal cost of $.002 = MC_i$, we find that the quasi-optimal prices are $p_1 = p_2 = .012$, and $p_3 = p_4 = .023$, which prices we have already shown involve a subsidy.

Since quasi-optimal prices depend on marginal costs and demand elasticities, whereas the constraints defining subsidy-free prices depend on the costs of alternative means of supply, it is no surprise that the two ideas are not necessarily compatible. Baumol-Bradford implicitly assume away the possibility of competitive entry, an assumption that is quite in keeping with their optimal taxation approach. Competitive entry, however, is the key idea of core constraints, and therein lies the crucial difference between the two results: the Baumol-Bradford optimal taxation approach is to maximize welfare subject to a break-even constraint, without regard to possible competitive entry or alternative means of supply (and

---

ing a joint cost, increasing returns enterprise into a constant returns business (e.g., the advent of freight transport by truck in competition with established rail transport), or vice versa. Although its difficulty and importance to public policy are clear, the problem is not treated in this paper.

[17] It should be borne in mind that the demand function to which we refer reflects the utility maximizing behavior of the consumer group, and is *not* the demand function seen by the water company. This latter function would become more elastic than the true demand function of the consumer group in the region where price was sufficiently high to make alternative means of supply attractive to the consumer. Thus, our demand elasticity measures consumer's attitudes toward the commodity *only*, and *not* the cost of alternative supply, or the degree of competition in the market.

hence, cross-subsidy). The cross-subsidy approach focuses on the costs of alternative supply to determine a set of prices, all of which provide positive incentives to forestall inefficient entry, and does not address directly the welfare maximization question.

If our public enterprise satisfies condition (1) so that it offers economies of joint production, are we assured that the core is never empty? Surprisingly and unfortunately, the answer is no; it appears that in a situation where average costs at first decrease and then increase, we may fail to obtain a core. Let us consider a new water company, located in the county next to our old water company. Our new company has three neighborhoods clustered about it, so all it need do to supply any one or combination of the three is to drill a well, build a storage tank, and construct the appropriate number of local distribution networks, the last item costing \$100 apiece, including pumping costs for 10,000 gallons/day/neighborhood. A geologist determines that a shallow well can be drilled to yield a maximum of 20,000 gallons/day at a cost of \$200 (including tank). If a higher output is desired, a deeper well must be drilled, at a cost of \$350. Thus, for $q_i = 10{,}000$, $i = 1, 2, 3$,

$$C(q_1, 0, 0) = C(0, q_2, 0) = C(0, 0, q_3) = \$300$$
$$C(q_1, q_2, 0) = C(q_1, 0, q_3) = C(0, q_2, q_3) = \$400$$
$$C(q_1, q_2, q_3) = \$650$$

since only the coalition of all three with its total demand of 30,000 gallons/day needs the more expensive well. Clearly, (1) is satisfied, but no core exists, since the core constraints are

$$\begin{aligned} r_1 + r_2 &\leqq 400 \\ r_1 + r_3 &\leqq 400 \\ r_2 + r_3 &\leqq 400 \\ 2(r_1 + r_2 + r_3) &\leqq 1200, \\ \text{or} \quad r_1 + r_2 + r_3 &\leqq 600 \end{aligned}$$

which cannot hold, since $r_1 + r_2 + r_3 = \$650$.

The existence of joint cost enterprises for which there are no prices free of subsidy could present serious difficulties in view of our discussion of market incentives. If no core exists and if entry into the market is free, then there will be no stable supply arrangement, since for any set of prices for any production and supply arrangement, there will always be an incentive for an entrepreneur to offer a new coalition lower prices in a new supply arrangement. Not even the coalition of all consumer groups can block this unstable behavior. In order to insure the stability of the most efficient cooperative solution, coercive intervention to restrict market entry becomes necessary.

III. The Cooperative Game with Interrelated Demands

In Section I, we considered the case of a multicommodity enterprise in which the demand for commodity i depended only upon the price of commodity i. In modeling the enterprise as a cooperative game, we assumed that demand and supply levels remained fixed, even under alternative supply arrangements. We then showed that if prices were subsidy-free in the "fixed demand" game, they remained so when prices and quantities changed under various supply arrangements provided that the cross-elasticities of demand are zero.

Unfortunately, no such simplifying result holds in the presence of cross-elasticities. For consider a firm which is about to offer a new service or commodity for sale; the new service will add to both the revenues and the costs of the firm. Our previous analysis suggested that if the added revenues exceeded the added costs, then the service was not being subsidized and in fact was a benefit to the other services. We therefore compared the revenues collected from this service with its incremental costs. However, in the presence of cross-elasticities, the revenues collected from the new service do not measure the added, or incremental revenues, since revenues from the other services may rise (if the service in question is a net complement) or decline (if the new service is a net substitute). If we consider varying prices, then the core will be smaller or larger than the fixed demand core, depending on whether the service is a substitute or complement for the remaining services.

To see this, we first define our "varying demand" game in so-called generalized characteristic function form (see Scarf).[18] Let

$$\pi(q^S) = R(q^S) - C(q^S)$$

where $q_i = q_i(p_1, \ldots, p_n)$.

$$\begin{aligned} (5) \qquad V^N &= \{p^N \mid \pi(q^N(p^N)) = 0\} \\ V^S &= \{p^S \mid \pi(q^S(p^S, p^{N-S})) \geqq 0 \\ &\qquad \text{for all } p^{N-S} \text{ for which} \\ &\qquad\qquad \pi(q^{N-S}(p^S, p^{N-S})) \geqq 0\} \end{aligned}$$

Clearly, V^N is the set of prices satisfying the zero profit constraint; V^S, however, needs explaining. In words, V^S is the set of prices for the coalition S which are feasible for S $(\pi(q^S) \geqq 0)$ for any feasible choice of prices for the coalition $N-S$. If the coalition S chooses prices in V^S, then it is not feasible for the coalition $N-S$ to choose prices which would cause negative profits for S. Thus, V^S represents prices the coalition S can obtain unilaterally, with no cooperation from members of $N-S$.

If the prices of the enterprise lay in the interior of V^S, for some $S \subset N$, then the coalition S could choose prices in V^S which were uniformly lower than the given prices, and furthermore unilaterally change the supply structure to realize the lower prices (since those prices are in V^S). Clearly, core prices should be in V^N but not V^S; hence,

$$(6) \qquad Core = V^N - \bigcup_S V^S$$

To illustrate what can occur in a game with nonzero cross-elasticities, we develop and extend a simple two-service example given by Zajac. This enables us to examine (a) the core of the game, (b) what happens if either service opts to leave (the "noncooperative" game), and (c) how the core of this game corresponds to the core of the fixed demand game.

Suppose we have a firm producing two commodities with \$10 in common fixed cost. Commodity 1 is produced with \$10 allocable fixed cost and a constant marginal of \$1. Commodity 2 requires \$60 of allocable fixed cost and a constant marginal cost of \$0.50. Thus

$$\begin{aligned} (7) \qquad C(q_1, 0) &= 20 + q_1 \\ C(0, q_2) &= 70 + .5q_2 \\ C(q_1, q_2) &= 80 + q_1 + .5q_2 \end{aligned}$$

The two goods are substitutes with linear demand functions over the relevant range given by:

$$\begin{aligned} (8) \qquad q_1 &= 100 - 10p_1 + 4p_2 \\ q_2 &= 40 + 4p_1 - 4p_2 \end{aligned}$$

where p_i is the price of commodity i. The profit function of the firm is denoted by $\pi(p_1, p_2)$ which is assumed constrained to zero, so that $V^{\{1,2\}}$ is the set of prices satisfying

$$\begin{aligned} (9) \qquad \pi(p_1, p_2) = &- 10p_1^2 + 8p_1p_2 - 4p_2^2 \\ &+ 108p_1 + 38p_2 - 200 = 0 \end{aligned}$$

This second-degree equation plots as an ellipse in (p_1, p_2) space. Only the lower left-hand portion of this ellipse is relevant, corresponding to our assumption that $\partial\pi/\partial p_i > 0$ (see footnote 14).

What happens if the two commodities are produced separately by two different firms? Each firm would have its own profit function π_i which would reflect the duplicated common costs, and would also depend upon the prices that the other firm would charge, because of the nonzero cross-elasticities. Hence,

---

[18] This approach was suggested to me by my colleague William Sharkey.

$$(10)\quad \begin{aligned} \pi_1(p_1, p_2) &= -10p_1^2 + 4p_1p_2 + 110p_1 \\ &\quad - 4p_2 - 120 = 0 \\ \pi_2(p_1, p_2) &= -4p_2^2 + 4p_1p_2 + 42p_2 \\ &\quad - 2p_1 - 70 = 0 \end{aligned}$$

In Figure 1, we show the curves $\pi=0$, $\pi_1=0$, and $\pi_2=0$ plotted in (p_1, p_2) space. It can be seen that

$$\begin{aligned} V^1 &= \{p_1 \mid p_1 \geqq p_1^* = \$1.204\} \\ V^2 &= \{p_2 \mid p_2 \geqq p_2^* = \$2.51\} \end{aligned}$$

In fact, p_1^* and p_2^* are the minimax prices of the noncooperative game, so that point $C=(p_1^*, p_2^*)$ is the solution to that game. Further, p_1^*, p_2^* is a solution to the simultaneous system (10), and the only one for which $\partial\pi_i/\partial p_i>0$.

The core of the game consists of the set of prices satisfying

$$\pi(p_1, p_2) = 0 \qquad p_1 \leqq p_1^* \qquad p_2 \leqq p_2^*$$

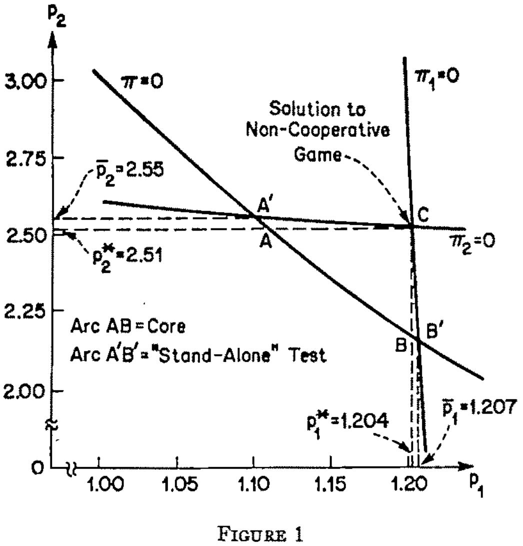


FIGURE 1

which is the arc AB in Figure 1.

Figure 1 also yields an interesting interpretation of the fixed demand (zero cross-elasticities) game. Suppose the prices lie on the curve $\pi=0$ but are to the right of $\pi_1=0$. Then $\pi_1>0$ and the revenues from commodity 1 exceed its stand-alone cost and thus do not lie in the core of the fixed demand game. Similarly, if the prices satisfied $\pi=0$ but lie above $\pi_2=0$, then the revenues from commodity 2 exceed its stand-alone cost, and are not in the core. The fixed demand core is the arc $A'B'$ in Figure 1, for which $p_1 \leqq \bar{p}_1 = \$1.207$ and $p_2 \leqq \bar{p}_2 = \$2.55$. This corroborates our previous assertion that the arc is smaller if we introduce positive cross-partials in the demand function. Thus, if the commodities in question are substitutes, there may be prices which pass an incremental cost test and yet involve cross-subsidy.

IV. Summary

In this paper, we have defined and analyzed the issues of cross-subsidization in enterprises with economies of joint production. We have couched the problem in game theoretic terms, which led us to note that not only individual commodities must be examined for possible cross-subsidization, but also groups of commodities ("coalitions"). We then explored the question of market incentives under subsidizing vs. subsidy-free prices, and noted that subsidization by prices in a free market would lead to competitive entry and instability of the joint enterprise, particularly when no subsidy-free prices existed. The discussion of the relationship between welfare-maximizing prices and subsidy-free prices suggested conditions under which these two criteria might clash, and the necessity of coercive intervention in restricting market entry if that were the case. This extension of a game theoretic analysis to the case of nonzero cross-elasticities indicates how the problem might be approached, and what difficulties might be encountered.

REFERENCES

E. P. Alexander, *Railway Practice*, New York 1887.

E. E. Bailey, *Economic Theory of Regulatory Constraint*, New York 1973.

W. J. Baumol and D. F. Bradford, "Optimal Departures from Marginal Cost Pricing," *Amer. Econ. Rev.*, June 1970, *60*, 265–83.

M. Boiteux, "Sur la gestion des Monopoles Publics astreints à l'équilibre budgetaire," *Econometrica*, Jan. 1956, *24*, 22–40.

K. Borch, "Application of Game Theory to Some Problems in Automobile Insurance," *ASTIN Bull.*, Sept. 1962, *2*, 210–21.

J. M. Buchanan, "An Economic Theory of Clubs," *Economica*, Feb. 1965, *32*, 1–14.

D. B. Gillies, "Some Theorems on N-Person Games," unpublished doctoral dissertation, Princeton Univ. 1953.

H. Hotelling, "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates," *Econometrica*, July 1938, *6*, 242–64.

A. E. Kahn, *The Economics of Regulation*, Vol. I, New York 1970.

S. C. Littlechild, "A Game Theoretic Approach to Public Utility Pricing," *Western Econ. J.*, June 1970, *8*, 162–66.

E. T. Loehman and A. B. Whinston, "A New Theory of Pricing and Decision-Making for Public Investment," *Bell J. Econ.*, Autumn 1971, *2*, 606–25.

——— and ———, "An Axiomatic Approach to Cost Allocations for Public Investment," *Publ. Finance Quart.*, 1974, *2*, 236–51.

——— and ———, "A Generalized Cost Allocation Scheme," presented at Conference on Externalities, Southern Illinois Univ., Feb. 1974.

R. D. Luce and H. Raiffa, *Games and Decisions, Introduction and Critical Survey*, New York 1957.

M. V. Pauly, "Clubs, Commonality, and the Core: An Integration of Game Theory and the Theory of Public Goods," *Economica*, Aug. 1967, *34*, 314–24.

———, "Cores and Clubs," *Publ. Choice*, Fall 1970, *9*, 53–65.

J. Quirk and R. Saposnik, *Introduction to General Equilibrium Theory and Welfare Economics*, New York 1968.

F. Ramsey, "A Contribution to the Theory of Taxation," *Econ. J.*, Mar. 1927, *37*, 47–61.

N. Ruggles, "Recent Developments in the Theory of Marginal Cost Pricing," *Rev. Econ. Stud.*, 1950, *17*, 107–26.

H. Scarf, "The Core of an N-Person Game," *Econometrica*, Jan. 1967, *35*, 50–69.

H. Trebing, "Competition: An Asset to Utility Regulation?," presented at Public Utilities Seminar of Amer. Marketing Assoc., Chicago 1967.

E. E. Zajac, "Some Preliminary Thoughts on Subsidization," presented at Conference on Telecommunications Research, Washington 1972.

Revealed Preference for Public Goods

By HAJIME HORI*

The most serious problem for the modern theory of optimal public expenditure is that although it has to be based on some social welfare function, individual preferences for public goods, which constitute the elements of any individualistic social welfare function, are hard to assess. It is argued that since every individual can freely participate in the consumption of a public good, there is an incentive for him to either underreport or overreport his preferences depending on how the reported preferences are used in the financing process of the public good.[1]

This argument will be correct if an individual's preferences for public goods can be expressed only directly. In many cases, however, the consumption of a public good has some technical or objective relationship with the consumption of some private goods. Thus, the benefit an individual derives from free highways depends on how much he spends on a car and gasoline. Even the benefit an individual derives from the judicial system depends on how much he spends to hire the services of lawyers. Also, an individual can enjoy the same security police provide by hiring private guards and installing locks, and so on. In these cases, if we know the technical relationships between the consumption of private goods and the consumption of public goods, then from the market data relating the amounts of private goods bought to prices and the data on the existing amounts of public goods, we might be able to infer (i) the amounts of *final goods* consumed (say the number of trips from one point to another, the degree of justice, the security from robbery, etc.), and (ii) their opportunity costs which the individual was willing to pay. But if we succeed in this inference, then all that remains to be done in order to find the individual's utility function will be to integrate back the relationship between the amounts of final goods consumed and their opportunity costs.

A trivial example can be provided by assuming that each public good is perfectly substitutable with a private good. Let y_i be the quantity of the ith public good and x_i be the quantity of a private good which is a perfect substitute for the ith public good. Then the benefit derived from the ith public good and its perfect substitute can be measured by x_i+y_i. Suppose we have ascertained an individual consumer's demand functions for private good when all the y_i are zero. We know from the revealed preference theory that if these demand functions satisfy a certain condition (namely the symmetry of the Slutsky matrix), then there exists a utility function unique up to a monotone transformation:

$$U = U(x_1, x_2, \ldots, x_n)$$

But since the ith private good is a perfect substitute for the ith public good, this consumer's utility function which covers the case where $y_i \neq 0$ is given by

---

* Assistant professor of economics, Tohoku University, Sendai, Japan. Much of the work for this paper was completed at Tulane University. I am indebted to an anonymous referee for useful suggestions and comments.

[1] See Richard Musgrave, Paul Samuelson (1964), and Hirofumi Shibata. The devices recently developed by Mordecai Kurz for determining the demand for public goods also accept this argument and try to establish environments in which there is no incentive for consumers to falsify their preferences.

$$V(x_1, \ldots, x_n; y_1, \ldots, y_n) = U(x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)$$

Admittedly this example is trivial. But, as will be shown below, a similar reasoning can be extended to more general situations. What is crucial is *the assumption that there exists a technical or objective relationship between the consumption of each public good and the consumption of some private goods.*

I

In order to formalize this assumption, suppose an individual's utility function is of the Strotz–Lancaster–Muth–Sandmo type,[2] and let its general representation be

$$(1) \qquad U = U(z) = U(z_1, z_2, \ldots, z_n)$$

where the z_i are the amounts of final goods consumed and $z=(z_1, \ldots, z_n)$. Each z_i is related to the amounts of private and public goods used according to the formula

$$(2) \qquad z_i = f^i(x^i; y_i), \qquad i = 1, \ldots, n$$

where x^i is the vector of the quantities of private goods and y_i is the quantity of a public good used for the production (within household) of the ith final good. Here we are assuming for simplicity that the production of each final good requires at most one public good but the extension to the case of more than one public good per final good is straightforward. *The household production technology* (or consumption technology) represented by the functions f^i, $i=1, \ldots, n$, is assumed to be common to every consumer. We also assume sufficient smoothness of the functions f^i.

Our purpose is to consider under what conditions we can deduce the specific form of the utility function (1), assuming that we know the objective household production technology (2) and the individual consumer's market behavior summarized by his demand functions for private goods,

$$(3) \qquad x_j = x_j(p, I, y)$$

where x_j is the quantity demanded of the jth private good, p is the price vector of private goods, I is income, and y is the vector of the supplies of public goods.

As was mentioned before, the main task is to find the marginal rate of substitution between each pair of final goods at each point in the final-goods space. But the specific exercises we have to go through for this purpose depend upon how private goods enter the household production of final goods. We will separately consider four possible forms of the household production technology, which differ from each other by the way private goods enter the household production process.

The first form of household production technology to be considered is *separable* in the sense that each private good enters the production of only one final good. Thus let

$$(4) \qquad z_i = f^i(x_{m_{i-1}+1}, x_{m_{i-1}+2}, \ldots, x_{m_i}; y_i), \qquad i = 1, \ldots, n$$

where $m_0=0$. There are assumed to be $m_n=m$ private goods, out of which $m_{i-1}+1, \ldots, m_i$ are used for the production of the ith final good. This form of household production function was considered by Strotz and Muth. We will take up this household production technology in Section II.

The second form to be considered is the one in which each private good is *divided* into multiple uses. For example, one's leisure time is divided into several distinct uses such as reading novels, camping in a national park, and so on. Let x_{ij} denote the amount of the jth private good used in the production of the ith final good. If there are m private goods, then we can write:

---

[2] See Robert Strotz, Kelvin Lancaster, Richard Muth, and Agnar Sandmo. It was Sandmo who first applied this type of utility function to the analysis of optimal public expenditure.

$$(5) \qquad z_i = f^i(x_{i1}, x_{i2}, \ldots, x_{im}; y_i), \qquad i = 1, \ldots, n$$

Although this can be regarded as a generalization of (4), a complication arises because each x_{ij} cannot usually be observed in the market; for each price-income-public goods combination, what we can observe in the market is the *total quantity* consumed of each private good, $x_j = \sum_{i=1}^{n} x_{ij}$, and not its decomposition x_{ij}. We will consider this form of household production technology in Sections III and IV.

The third form of the household production technology is that of *assumed indivisibility* (or joint production). Namely, the use of private goods $x_1, \ldots, x_m$ in the production of one final good necessarily contributes to the production of other final goods and it is impossible to measure how much of x_1, say, goes into the production of z_1 or z_2 or z_3 and so on. Thus this form of household production technology can be represented by

$$(6) \qquad z_i = f^i(x_1, x_2, \ldots, x_m; y_i), \qquad i = 1, \ldots, n$$

This can be regarded as another generalization of (4). Some properties of consumer behavior which incorporate this type of household production technology were considered by Lancaster under an additional assumption that the functions f^i are linear.[3] We will consider this household production technology in Sections V and VI.

The final form to be considered is a mixture of (5) and (6) and can be written as

$$(7) \quad z_i = f^i(x_{i1}, \ldots, x_{im}; x_{m+1}, \ldots, x_M; y_i), \qquad i = 1, \ldots, n$$

where each of the private goods 1 through m is divided into various distinct uses while each of the private goods $m+1$ through M enters the production of various final goods jointly. We will consider this household production technology in Sections VII and VIII.

II

Let us start with the simplest case where each private good enters the household production of only one final good:

$$(4) \quad z_i = f^i(x_{m_{i-1}+1}, \ldots, x_{m_i}; y_i), \qquad i = 1, \ldots, n, \ m_0 = 0, \ m_n = m$$

In this case we can in principle observe m distinct demand functions

$$(8) \quad x_j = x_j(p, I, y), \quad j = 1, \ldots, m$$

where $p = (p_1, \ldots, p_{m_1}, p_{m_1+1}, \ldots, p_m)$. By substituting (8) in (4), we can find what z_i our consumer chooses for each combination of prices, income, and public goods:

$$(9) \quad \bar{z}_i = f^i(x(p, I, y)), \quad i = 1, \ldots, n$$

In order to evaluate the marginal rates of substitution between final goods, we first define the cost function for each final good by

$$(10) \quad C_i(z_i; p, y_i) = \min \left\{ \sum_{j=m_{i-1}+1}^{m_i} p_j x_j \;\middle|\; f^i(x_{m_{i-1}+1}, \ldots, x_{m_i}; y_i) = z_i \right\}, \qquad i = 1, \ldots, n$$

Note that these cost functions are common to every individual. The budget constraint for an individual with income I, defined in the final-goods space, is given by

$$(11) \qquad \sum_{i=1}^{n} C_i(z_i; p, y_i) = I$$

A rational consumer will maximize his utility subject to the budget constraint (11). Thus the combination of final goods he chooses, given by (9), satisfies (11).

---

[3] The linearity of the functions f^i, however, poses a problem to our approach, as will be shown in Section V.

Moreover, at point $\bar{z}=(\bar{z}_1, \ldots, \bar{z}_n)$ given by (9), the budget surface must be tangent to one of his indifference surfaces. Thus, if we let $R_{ii'}(z)$ denote the marginal rate of substitution between the ith and the i'th final goods at point z, we obtain

$$R_{ii'}(\bar{z}) = -\frac{C_i'(\bar{z}_i)}{C_{i'}'(\bar{z}_{i'})}, \quad i, i' = 1, \ldots, n, \quad i \neq i' \tag{12}$$

If the range of the demand functions for final goods (9) is equal to the whole final-goods space, then we can obtain the marginal rates of substitution for all the points from (12).

III

Next we consider the case where each private good is divided into multiple uses:

$$z_i = f^i(x_{i1}, x_{i2}, \ldots, x_{im}; y_i), \quad i = 1, \ldots, n \tag{5}$$

Analogously to equations (10) and (11) in the previous section, we can find the cost functions and the budget constraint from (5). But in the present case, this is not enough for finding the marginal rates of substitution. Since the available market data in the present case take the form of m demand functions

$$\sum_{i=1}^{n} x_{ij} \equiv x_j = x_j(p, I; y), \quad j = 1, \ldots, m \tag{13}$$

which relate p, I, and y to the total quantities x_j and not to their decompositions x_{ij}, we cannot directly relate p, I, and y to z, and therefore we cannot directly pinpoint at what point in the final-goods space the budget surface is tangent to an indifference surface.

In order to find the z which the consumer has in mind when he buys $x=(x_1, \ldots, x_m)$ according to (13), let us consider the problem:

$$\text{Minimize } \sum_{j=1}^{m} p_j x_j \text{ over } x \tag{14}$$

subject to $f^i(x_{i1}, \ldots, x_{im}; y_i) = z_i$, $i = 1, \ldots, n$,

$$\sum_{i=1}^{n} x_{ij} = x_j, \quad j = 1, \ldots, m$$

for each z. This is an objective part of our consumer's utility maximization and has a unique solution

$$x_{ij} = g_{ij}(z_i; p, y_i), \quad i = 1, \ldots, n, \quad j = 1, \ldots, m \tag{15}$$

$$x_j = g_j(z; p, y) \equiv \sum_{i=1}^{n} g_{ij}(z_i; p, y_i), \quad j = 1, \ldots, m$$

if the household production functions f^i are concave-contoured (i.e., if their isoquants are convex toward the origin) in the m-dimensional private-goods space.

Equations (15) immediately give the cost functions

$$C_i(z_i; p, y_i) \equiv \sum_{j=1}^{m} p_j g_{ij}(z_i; p, y_i), \quad i = 1, \ldots, n \tag{16}$$

and the budget constraint by (11) and (16). At the same time, note that since the demand functions (13) are the result of the consumer's utility maximization, they must be consistent with (15). Thus the z which the consumer has in mind when he buys $x=(x_1, \ldots, x_m)$ according to (13) must satisfy

$$g_j(z; p, y) = x_j(p, I; y), \quad j = 1, \ldots, m \tag{17}$$

for each combination of p, I, and y. In (17), we have m equations for n unknowns z_i and the preference orderings of the consumer appear as the observable demand functions on the right-hand side. This system of equations has at least one solution z if the consumer is consistent. But for observers, the crucial question is the uniqueness of a solution. If (17) has multiple solutions, then we cannot tell at what z the budget surface (11) is tangent to an indifference surface and therefore there is no way of finding the structure of the indifference map. If a solution is unique for a sufficient number of combinations of p, I, and y so that we can cover the whole final-goods space by varying p, I, and y in

$$z_i = z_i(p, I, y), \qquad i = 1, \dots, n \tag{18}$$

then, in the same way as in the previous section, we will be able to arrive at the marginal rates of substitution between final goods as functions of z, given by (12).

There is another way of formulating the problem. For each $x = (x_1, \dots, x_m)$, consider

$$\text{(19)} \quad \text{Maximize } \sum_{i=1}^{n} C_i(z_i; p, y_i) \text{ over } z$$

$$\text{subject to } z_i = f^i(x_{i1}, \dots, x_{im}; y_i), \qquad i = 1, \dots, n$$

$$\sum_{i=1}^{n} x_{ij} = x_j, \qquad j = 1, \dots, m$$

where the functions $C_i(z_i; p, y_i)$ are defined by (16). Obviously (19) is a dual to the cost minimization problem (14). Therefore the uniqueness of a solution to (17) is equivalent to the uniqueness of a solution $z_i = z_i(p, I, y)$ of (19) for each $x = x(p, I, y)$ given by (13).

We have formulated the problem of finding the chosen z in two equivalent ways, namely as the problem of the uniqueness of a solution to either (17) or (19). The first formulation is probably better suited to formal analysis while the second formulation admits of a geometrical analysis.

Equations (17) immediately show that the uniqueness requires $m \geqq n$. The case where $m > n$ can be reduced to the case of $m = n$ because the rationality of the consumer implies the existence of a solution to (17) and therefore all we have to do is to pick a suitable set of n equations and consider the uniqueness for it. For the case where $m = n$, the problem of the uniqueness of a solution to equations similar to (17) has been considered in a different context and some sufficient conditions are known. Interested readers are referred to David Gale and H. Nikaido, and Nikaido.[4]

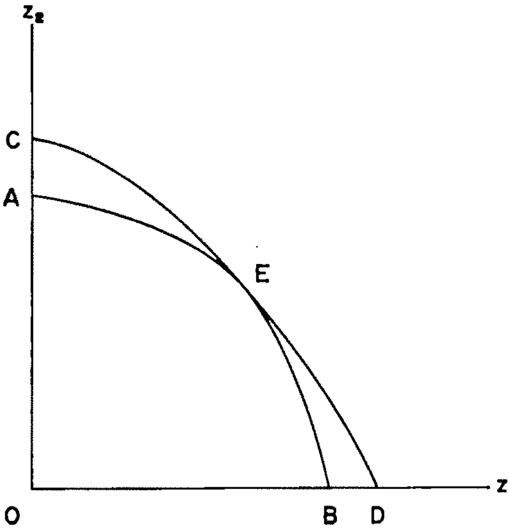


FIGURE 1

Figures 1 and 2 illustrate the uniqueness problem as formulated in (19). The curves AEB in Figure 1 and $AEE'B$ in Figure 2 are the possibility frontier; namely the frontier of the set of all the z satisfying

---

[4] If the $f^i(x_{i1}, \dots, x_{im}; y_i)$, $i = 1, \dots, n$, are linearly homogeneous in private goods $x_{i1}, \dots, x_{im}$, then the most refined sufficient condition for the uniqueness of a solution, developed by I. F. Pearce in the context of the factor-price-equalization theorem, can be utilized. If we let $x^i(p; y_i)$ denote the m-dimensional vector of the minimum-unit-cost inputs for the ith final good, then (17) or (19) has a unique solution if the n vectors $x^i(p; y_i)$ are linearly independent for each p. See Pearce (1959, 1967) and Samuelson (1967). In the present context, however, we cannot usually expect the linear homogeneity in private goods alone of the household production functions. The sufficient conditions developed by Gale and Nikaido, and Nikaido do not depend on the homogeneity assumption.

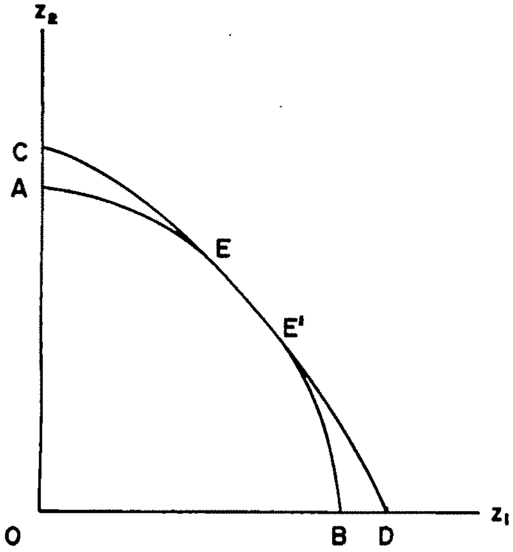


FIGURE 2

$$z_i = f^i(x_{i1}, \ldots, x_{im}; y_i), \qquad i = 1, 2$$
$$x_{1j} + x_{2j} = x_j, \qquad j = 1, \ldots, m$$

for the x actually bought at some p, I, and y. The curves *CED* in Figure 1 and *CEE'D* in Figure 2 are the budget curve that corresponds to the maximum total cost obtainable on the possibility frontier when the prices of private goods and the supplies of public goods are p and y. A rational consumer will choose the z at which the budget curve is tangent to an indifference curve. At this z the possibility frontier will also be tangent to the same indifference curve because, in addition to choosing an x, the consumer has to divide each x_j into x_{1j} and x_{2j} so as to maximize his utility. Thus at the z which the consumer has in mind in buying an x, the budget curve and the possibility frontier are tangent to each other.

Figure 1 depicts the case of a unique solution and Figure 2 the case of multiple solutions. In Figure 2, we know that an indifference curve must be tangent to both the budget curve and the possibility frontier at some point between E and E'. But from our data we cannot tell which point it is.

IV

Since the uniqueness of a solution to (17) or (19) is crucial, it will be worthwhile to consider this problem further, using a simple example. In this section, we will assume that the household production functions are of the Cobb-Douglas type.

Let

$$(20) \qquad z_i = y_i^{b_i} \prod_{j=1}^{m} x_{ij}^{a_{ij}}, \qquad i = 1, \ldots, n$$

$$\text{where} \quad 0 \leq a_{ij}, \qquad i = 1, \ldots, n, \quad j = 1, \ldots, m$$
$$0 < \sum_{j=1}^{m} a_{ij}, \qquad i = 1, \ldots, n$$
$$0 \leq b_i, \qquad i = 1, \ldots, n$$

In this case, equations (17) become

$$(21) \qquad \sum_{i=1}^{n} h_{ij} z_i^{1/a_i} = x_j(p, I; y), \qquad j = 1, \ldots, m$$

where

$$h_{ij} = y_i^{-bi/a_i} \frac{a_{ij}}{p_j} \prod_{l=1}^{m} \left(\frac{p_l}{a_{il}}\right)^{a_{il}/a_i}$$
$$i = 1, \ldots, n, \quad j = 1, \ldots, m$$
$$a_i = \sum_{j=1}^{m} a_{ij}, \qquad i = 1, \ldots, n$$

Assume that $m = n$. Then a solution to (21) is unique if and only if

$$\begin{vmatrix} h_{11} & \ldots & h_{n1} \\ \vdots & & \vdots \\ h_{1n} & & h_{nn} \end{vmatrix} \neq 0$$

But it is easy to check that

$$\begin{vmatrix} h_{11} & \ldots & h_{n1} \\ \vdots & & \vdots \\ h_{1n} & \ldots & h_{nn} \end{vmatrix} = \prod_{i=1}^{n} w_i \prod_{j=1}^{n} \frac{1}{p_j} \begin{vmatrix} a_{11} & \ldots & a_{n1} \\ \vdots & & \vdots \\ a_{1n} & \ldots & a_{nn} \end{vmatrix}$$

where

$$w_i = y_i^{-b_i/a_i} \prod_{j=1}^{n} \left(\frac{p_i}{a_{ij}}\right)^{a_{ij}/a_i}, \quad i = 1, \ldots, n$$

Therefore the system of equations (21) has a unique solution if and only if $|a_{ij}| \neq 0$. Including the case where $m>n$, we can conclude that (21) has a unique solution if and only if the n vectors of constants

$$\begin{bmatrix} a_{i1} \\ a_{i2} \\ \cdot \\ \cdot \\ \cdot \\ a_{im} \end{bmatrix}, \quad i = 1, \ldots, n$$

are linearly independent.

V

The third form of the household production technology to be considered is that of *assumed indivisibility* (or joint production). The production of a final good with the use of private goods $x_1, \ldots, x_m$ (and the public good y_i) is necessarily accompanied by the production of other final goods:

$$(6) \qquad z_i = f^i(x_1, x_2, \ldots, x_m; y_i), \quad i = 1, \ldots, n$$

Unlike the second case considered above, to find the combination of final goods chosen by an individual poses no problem, because all the arguments of each household production function are observable variables. Let

$$(22) \quad B(p, I, y) = \left\{ z \,\middle|\, z_i \leq f^i(x_1, \ldots, x_m; y_i), \; i=1, \ldots, n, \; \sum_{j=1}^{m} p_j x_j = I \right\}$$

$B(p, I, y)$ is the set of all the z the individual can obtain when the prices of private goods, income, and the quantities of public goods are p, I, and y. The frontier of this set gives the budget surface on which the chosen z has to lie. Let

$$(23) \qquad F(z_1, \ldots, z_n; p, I, y) = 0$$

denote this surface. If $\bar{x}$ is the chosen vector of private goods, we can find the chosen vector of final goods, denoted by $\bar{z}$, from equations (6). Then one of the indifference surfaces must be tangent to (23) at $\bar{z}$. Thus *if the budget surface* (23) *is differentiable*, then the marginal rate of substitution between the ith and i'th final goods at $\bar{z}$ is given by

$$(24) \quad R_{i,i'}(\bar{z}) = -\frac{\partial F(\bar{z}; p, I, y)/\partial z_i}{\partial F(\bar{z}; p, I, y)/\partial z_{i'}}, \quad i, i' = 1, \ldots, n, \quad i \neq i'$$

The function F, however, is not necessarily differentiable, even if the functions f^i are differentiable, as shown in Figures 3 to 5. Figure 3 depicts the case where there are two final goods but only one private good. As the amount of the private good is increased, the obtainable combination of z_1 and z_2 moves along the curve *ODB*. Given the price of the private good, income, and the amounts of public goods, there is only one relevant combination of z_1 and z_2, represented by point *B*. The

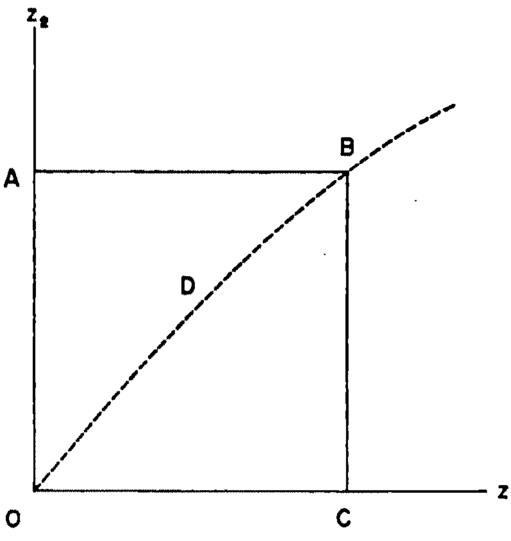


FIGURE 3

budget curve is ABC, and any rational individual will choose point B. Utility maximization at point B is consistent with an infinite variety of marginal rates of substitution. This situation is apparently due to the fact that the consumer does not have a choice among different combinations of z_1 and z_2 since there is only one choice variable. Generalization to the case of more final goods and more private goods is immediate. Thus, in order to avoid this situation, we have to require that

$$m \geq n \tag{25}$$

Figure 4 represents the case where the household production technology is linear in x_1, x_2, and y:

$$z_1 = b_{11}x_1 + b_{12}x_2 + c_1y_1$$
$$z_2 = b_{21}x_1 + b_{22}x_2 + c_2y_2$$

The ray EB starting at E is the locus of all the z obtainable from private good 1 alone, and the ray EC starting at E is the locus of all the z obtainable from private good 2 alone. Thus points B and C denote the combinations of z_1 and z_2 obtainable when all the income is spent on private goods 1 and 2, respectively. The budget surface $ABCD$ has kinks at B and C because of the nonnegativity constraint $x_j \geq 0$, $j = 1, 2$. Again in this case, if the chosen combination of x_1 and x_2 corresponds to point B or point C (which is quite likely), we cannot infer the marginal rate of substitution between final goods 1 and 2.

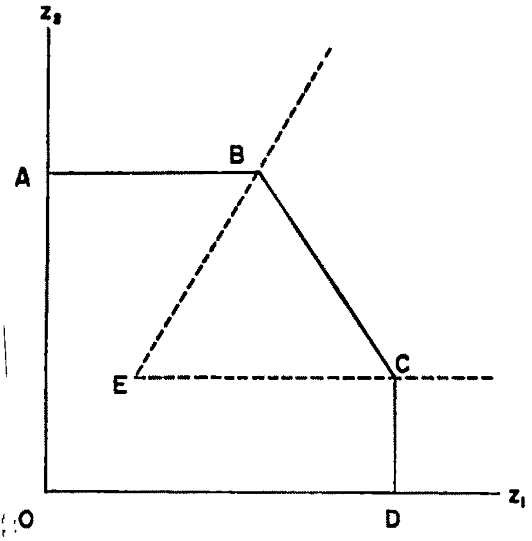


Figure 4

In Figure 4 the kinks of the budget curve are due to the *effective* nonnegativity constraint. Thus, for each $i = 1, \ldots, n$, let

$$J_i = \left\{ j \,\middle|\, \frac{\partial f^i}{\partial x_j} \neq 0,\ x_j = 0 \right\}$$

$$v^i = (v^i_1, \ldots, v^i_m), \qquad v^i_j > 0 \qquad \text{for } j \in J_i$$

If the household production functions have the property that

$$\lim_{h\to 0} \frac{d}{dh} f^i(x + hv^i; y_i) = \lim_{h\to 0} \sum_{j=1}^{m} \frac{\partial}{\partial x_j} f^i(x + hv^i; y_i)\cdot v^i_j = \infty, \qquad i = 1, \ldots, n \tag{26}$$

then the budget surface will not show this type of kink.

There is one more source of nondifferentiability. Suppose $n = m = 2$, and for each $i = 1, 2$, consider

$$\text{Maximize } z_i = f^i(x_1, x_2, y_i)$$

$$\text{subject to} \qquad p_1x_1 + p_2x_2 = I$$

Let $x^i = (x^i_1, x^i_2)$ denote the optimum vector for the ith maximization problem, $i = 1, 2$, and let

$$z^i_1 = f^1(x^i; y_i)$$
$$z^i_2 = f^2(x^i; y_i)$$
$$z^i = (z^i_1, z^i_2), \qquad i = 1, 2$$

If $z^1 = z^2$, then there is no tradeoff between the productions of final goods 1 and 2, as is shown in Figure 5. In Figure 5 as x_1 is increased from zero and x_2 is decreased from I/p_2 subject to $p_1x_1 + p_2x_2 = I$, both z_1

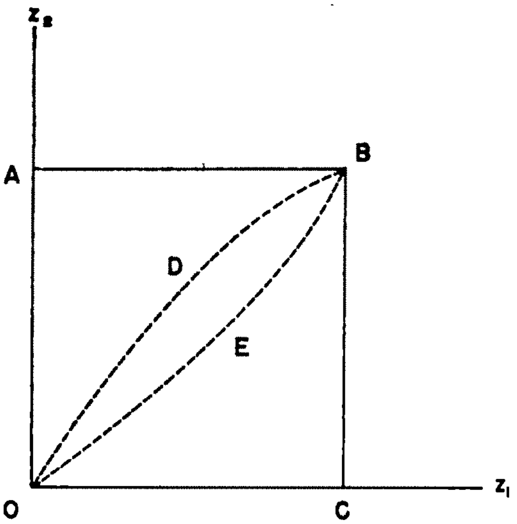


FIGURE 5

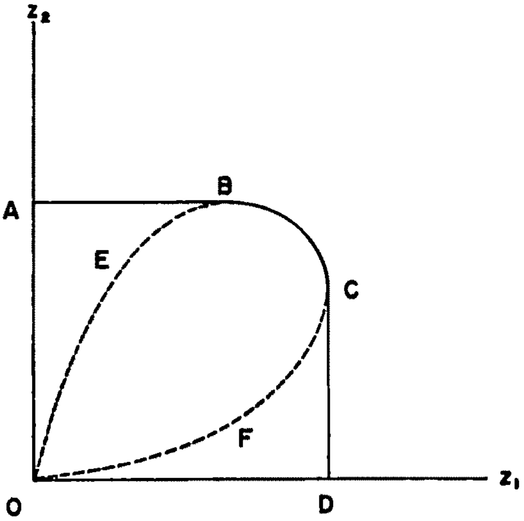


FIGURE 6

and z_2 increase along the curve *ODB* and reach their maxima simultaneously at point *B*. As x_1 is further increased and x_2 is further decreased, both z_1 and z_2 decrease. As a result, the budget curve is given by *ABC* with a kink at *B*. In this case, every rational consumer will choose point *B*, and again it is impossible to find the marginal rate of substitution between z_1 and z_2.

Figure 6 gives a smooth budget curve which results if $z^1 \neq z^2$. Point *B* corresponds to vector z^2 and point *C* corresponds to vector z^1. As x_1 is increased from zero and x_2 decreased from I/p_2, both z_1 and z_2 increase and z_2 reaches its maximum at point *B*. If x_1 is further increased and x_2 further decreased, z_2 decreases but z_1 increases until z_1 reaches its maximum at point *C*. After that, both z_1 and z_2 decrease along *CFO*.

In order to have a smooth budget surface in the general case of n final goods and m private goods with $m \geq n$, it will be necessary (and sufficient together with (25) and (26)) to have a smooth budget curve like Figure 6 on each 2-dimensional slice of the n-dimensional final-goods space. Thus, for each pair of final goods i_1 and i_2, define an $(n-2)$-dimensional vector $z^{i_1 i_2}$ by

$$z^{i_1 i_2} = (z_1, z_2, \dots, z_{i_1-1}, z_{i_1+1}, \dots, z_{i_2-1}, z_{i_2+1}, \dots, z_n)$$

For $i = i_1$ and $i = i_2$, consider

$$\text{(27)} \quad \text{Maximize} \quad z_i = f^i(x_1, \dots, x_m; y_i)$$

$$\text{subject to} \qquad \sum_{j=1}^{m} p_j x_j = I,$$

$$f^k(x_1, \dots, x_m; y_k) = z_k, \qquad k \neq i_1, i_2$$

Let x^i denote the optimum vector of x for the ith maximization problem and let

$$\overset{i}{z}_k = f^k(\overset{i}{x}; y_k)$$

$$\overset{i}{z} = (\overset{i}{z}_1, \overset{i}{z}_2, \dots, \overset{i}{z}_n), \qquad k = 1, \dots, n; \quad i = i_1, i_2$$

Both z^{i_1} and z^{i_2} are functions of $z^{i_1 i_2}$ (and p, I^*, y). Thus the condition for the smoothness of the budget surface for given p, I, and y is given by

$$\text{(28)} \quad z^{i_1} \neq z^{i_2} \quad \text{for all } z^{i_1 i_2}, i_1, i_2 = 1, \dots, n, \quad i_1 \neq i_2$$

Conditions (25), (26), and (28) constitute sufficient conditions for the existence of a smooth budget surface.

VI

As an example, suppose that the household production functions are of the Cobb-Douglas type:

$$(29)\quad f^i(x_1, \ldots, x_m; y_i) = y_i^{b_i} \prod_{j=1}^{m} x_j^{a_{ij}}, \qquad i = 1, \ldots, n$$

$$\text{where } 0 \leq a_{ij} < 1, \qquad i = 1, \ldots, n, \quad j = 1, \ldots, m$$

$$0 \leq b_i, \qquad i = 1, \ldots, n$$

$$\text{and} \quad n \leq m$$

The conditions (25) and (26) are satisfied. We will show that condition (28) is satisfied and the budget surface is smooth if

$$(30)\quad a^i = (a_{i1}, \ldots, a_{im}), \qquad i = 1, \ldots, n$$

are linearly independent.

In the following, we will take $i_1 = n-1$, $i_2 = n$. Also, we will number the goods and the constants in such a way that

$$(31)\quad \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} \neq 0$$

$$\Delta \equiv \begin{vmatrix} a_{11} & \cdots & a_{1,n-2} \\ \vdots & & \vdots \\ a_{n-2,1} & \cdots & a_{n-2,n-2} \end{vmatrix} \neq 0$$

Now the constrained maximization problems to determine x^{n-1}, z^{n-1}, x^n, and z^n are

$$\text{Maximize} \quad z_i = \prod_{j=1}^{m} x_j^{a_{ij}} y_i^{b_i}, \qquad i = n-1, n$$

$$\text{subject to} \quad \prod_{j=1}^{m} x_j^{a_{ij}} y_i^{b_i} = \bar{z}_i, \qquad i = 1, \ldots, n-2$$

$$\sum_{j=1}^{m} p_j x_j = I$$

For each of $i = n-1$, n, the first-order maximum conditions can be written as

$$(32)\quad \frac{p_k x_k \Delta - \sum_{j=1}^{n-2} \sum_{l=1}^{n-2} p_j x_j a_{lk} \Delta_{lj}}{p_{n-1} x_{n-1} \Delta - \sum_{j=1}^{n-2} \sum_{l=1}^{n-2} p_j x_j a_{l,n-1} \Delta_{lj}} = \frac{a_{ik}\Delta - \sum_{j=1}^{n-2} \sum_{l=1}^{n-2} a_{ij} a_{lk} \Delta_{lj}}{a_{i,n-1}\Delta - \sum_{j=1}^{n-2} \sum_{l=1}^{n-2} a_{ij} a_{l,n-1} \Delta_{lj}},$$

$$k = n, n+1, \ldots, m$$

where Δ_{lj} is the cofactor of a_{lj} in Δ.

Suppose $z^{n-1} = z^n$. Then there exists a vector $x = (x_1, \ldots, x_m)$ which satisfies the $(m-n+1)$ equations in (32) for both $i = n-1$ and $i = n$. Therefore, for each k, the right-hand side of (32) must have the same value for both $i = n-1$ and $i = n$. In particular, by taking $k = n$, we must have

$$\begin{aligned} (33)\quad B \equiv & \left(a_{nn}\Delta - \sum_{j=1}^{n-2} \sum_{l=1}^{n-2} a_{nj} a_{ln} \Delta_{lj}\right) \\ & \cdot \left(a_{n-1,n-1}\Delta - \sum_{j=1}^{n-2} \sum_{l=1}^{n-2} a_{n-1,j} a_{l,n-1} \Delta_{lj}\right) \\ & - \left(a_{n,n-1}\Delta - \sum_{j=1}^{n-2} \sum_{l=1}^{n-2} a_{nj} a_{l,n-1} \Delta_{lj}\right) \\ & \cdot \left(a_{n-1,n}\Delta - \sum_{j=1}^{n-2} \sum_{l=1}^{n-2} a_{n-1,j} a_{ln} \Delta_{lj}\right) = 0 \end{aligned}$$

But it can be shown by using Sylvester's determinant identity that

$$(34)\quad B = \Delta \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & & a_{nn} \end{vmatrix}$$

Therefore, by (31), B cannot vanish. This completes the proof that if (30) holds then $z^{i_1} \neq z^{i_2}$, and therefore the budget surface is smooth.

VII

Finally let us consider a household production technology which is a mixture of the forms considered above:

$$(7) \quad z_i = f^i(x_{i1}, \ldots, x_{im}; x_{m+1}, \ldots, x_M; y_i), \quad i = 1, \ldots, n$$

where each of the private goods 1 through m is divided into different uses while each of the private goods $m+1$ through M enters the production of various final goods indivisibly. For each given p, I, and y, the market data we can use are the M quantities x_j, $j=1, \ldots, m, m+1, \ldots, M$, where

$$x_j = \sum_{i=1}^{n} x_{ij} \quad \text{for} \quad j = 1, \ldots, m$$

Thus, both of the problems we considered before are present: namely, (i) the nonsmoothness of the budget surface and (ii) the impossibility of finding the chosen vector z of final goods.

(i) As before, we will assume (26) in order to exclude the kinks of the budget surface arising from corner solutions. In the case of pure indivisibility, possible lack of tradeoffs between different final goods was another source of nonsmoothness. In the present case, however, if the household production of each final good requires at least one private good which is not used jointly, namely if

$$(35) \quad \frac{\partial f^i}{\partial x_{i,j_i}} \not\equiv 0 \quad \text{for some } j_i,\ 1 \leq j_i \leq m, \quad \text{and for each } i = 1, \ldots, n$$

then along the budget surface there always exists a tradeoff between each pair of final goods, and therefore the second source of nondifferentiability disappears. In fact, let $C(z)$ denote the minimum cost of acquiring z. Thus $C(z)$ is the solution to the problem:

$$(36) \quad \text{Minimize} \sum_{j=1}^{m} \sum_{i=1}^{n} p_j x_{ij} + \sum_{j=m+1}^{M} p_j x_j$$

$$\text{subject to } f^i(x_{i1}, \ldots, x_{im}; x_{m+1}, \ldots, x_M; y_i) = z_i, \quad i = 1, \ldots, n$$

Under assumptions (26) and (35), there exists a j_i for each i such that one of the necessary conditions for (36) is

$$p_{j_i} - \lambda_i \frac{\partial f^i}{\partial x_{i,j_i}} = 0$$

where λ_i is the Lagrangean multiplier for the ith constraint. As usual, λ_i gives the marginal cost of the ith final good and thus

$$\frac{\partial C}{\partial z_i} = \lambda_i = \frac{p_{j_i}}{\partial f^i / \partial x_{i,j_i}}, \quad i = 1, \ldots, n$$

are well defined and positive. Thus the budget surface

$$C(z) = I$$

is differentiable and its slope satisfies

$$\left. \frac{dz_i}{dz_{i'}} \right|_{C(z)=I} = -\frac{\partial C/\partial z_{i'}}{\partial C/\partial z_i} < 0, \quad i, i' = 1, \ldots, n, \quad i \neq i'$$

(ii) The problem of finding the chosen z takes a form slightly different from the one considered in Sections III and IV. Due to the assumed indivisibility, the minimization problem (36) may not give unique solutions for x_j, $j=m+1, \ldots, M$, even if the f^i have convex isoquants. In its stead, however, we can use the fact that the market data directly give the value of x_j, $j=m+1, \ldots, M$, for each p, I, and y. Let

$$\bar{x} = (x_{m+1}, \ldots, x_M)$$

For each $\bar{x}$ and z, consider the minimization problem (36) with respect to x_{ij}, $i=1, \ldots, n$, $j=1, \ldots, m$. If the f^i are concave-contoured in private goods, this will give a unique optimum solution

$$x_{ij} = g_{ij}(\bar{x}, z_i; p, y_i), \quad i = 1, \dots, n, \quad j = 1, \dots, m \tag{37}$$

Substitute (37) in (36) and perform the minimization, this time with respect to $\bar{x}$. This will give $M-m$ implicit equations

$$g_j(\bar{x}, z; p, y) = 0, \ j = m+1, \dots, M \tag{38}$$

The rationality of the consumer implies the consistency between (37) and (38) on the one hand and the M demand functions on the other in the following form:

$$\sum_{i=1}^{n} g_{ij}(\bar{x}(p, I; y), z_i; p, y_i) = x_j(p, I, y), \quad j = 1, \dots, m$$
$$g_j(\bar{x}(p, I; y), z; p, y) = 0, j = m+1, \dots, M \tag{39}$$

where $\bar{x}(p, I; y)$ is the vector of $M-m$ demand functions $x_j(p, I; y), j = m+1, \dots, M$. Thus the problem of finding the chosen z reduces to the unique solvability of (39) for z.

As in Section III, the uniqueness of an optimum z of the dual maximization problem

(40) Maximize

$$C(z; p, y) = \min\left\{ \sum_{j=1}^{m} \sum_{i=1}^{n} p_j x_{ji} + \sum_{j=m+1}^{M} p_j x_j \,\middle|\, f^i(x_{i1}, \dots, x_{im}, x_{m+1}, \dots, x_M; y_i) = z_i, \quad i = 1, \dots, n \right\}$$

over z, subject to

$$z_i = f^i(x_{i1}, \dots, x_{im}, \bar{x}(p, I; y); y_i), \quad i = 1, \dots, n$$

$$\sum_{i=1}^{n} x_{ij} = x_j(p, I; y), \quad j = 1, \dots, m$$

is equivalent to the unique solvability of (39).

VIII

As an example, let us consider

$$z_i = \prod_{j=1}^{m} x_{ij}^{a_{ij}} \prod_{j=m+1}^{M} x_j^{a_{ij}} y_i^{b_i}, \ i = 1, \dots, n \tag{41}$$

where $a_{ij} \geqq 0$. For simplicity we assume that $a_{i1} > 0$ for $i = 1, \dots, n$. We will show that the linear independence of the n vectors

$$\begin{pmatrix} a_{i1} \\ a_{i2} \\ \cdot \\ \cdot \\ a_{iM} \end{pmatrix}, \quad i = 1, \dots, n$$

is necessary and sufficient for the unique solvability of (39). As before we can safely assume that $M = n$.

The first-order conditions for the minimization problem (36) under the assumption that the f^i are given by (41) can be arranged to yield

$$x_{ij} = \frac{p_1}{p_j} \frac{a_{ij}}{a_{i1}} x_{i1}, \qquad i = 1, \dots, n, \quad j = 1, \dots, m$$
$$x_j = \frac{p_1}{p_j} \sum_{i=1}^{n} \frac{a_{ij}}{a_{i1}} x_{i1}, \quad j = m+1, \dots, n \tag{42}$$

Let $\bar{x}_1, \dots, \bar{x}_n$ be the quantities of the n private goods bought. Then the x_{ij} and the x_j which correspond to the chosen z satisfy (39). Using (42), we obtain

$$\begin{pmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & & a_{n2} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{1n} & a_{2n} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} x_{11}/a_{11} \\ x_{21}/a_{21} \\ \cdot \\ \cdot \\ \cdot \\ x_{n1}/a_{n1} \end{pmatrix} = \begin{pmatrix} \bar{x}_1 \\ p_2\bar{x}_2/p_1 \\ \cdot \\ \cdot \\ \cdot \\ p_n\bar{x}_n/p_1 \end{pmatrix} \tag{43}$$

By assumption the coefficient matrix is nonsingular and therefore the x_{i1} are uniquely determined by the $\bar{x}_j$. Substituting the results in (42), we find the x_{ij} and x_j. Then from (41) we can find the z that uniquely corresponds to $\bar{x}$. Since equations (43) are linear in the x_{i1}, the linear independence of the n constant vectors is not only sufficient but also necessary.

IX

In order to see whether or not a consumer can be regarded to be revealing his preferences for public goods through his market behavior, we have considered in Sections II through VIII whether we can infer the marginal rates of substitution among final goods based on our knowledge concerning the household production technology and the demand functions for private goods for the following four types of household production technology:

(i) each private good is used in the production of only one final good;
(ii) each private good is divided into multiple distinct uses;
(iii) each private good contributes to the production of multiple final goods indivisibly;
(iv) mixture of (ii) and (iii).

Major results of our analysis may be summarized in the following way:

In case (i), sufficient smoothness of the household production functions are enough for the inference.

In case (ii), if the household production functions are concave-contoured in the private-goods space, then the problem reduces to the uniqueness of a solution to the system of equations (17) to which we can apply some sufficient conditions developed elsewhere. Or alternatively, the problem reduces to the uniqueness of the maximization problem (19).

In case (iii), the problem reduces to the differentiability of the budget surface defined in the final-goods space. Conditions (25), (26), and (28) are sufficient for this purpose.

In case (iv), if the production of each final good requires at least one private good which is not used in an indivisible fashion, and if the household production functions are concave-contoured in the private-goods space, then the problem reduces to the uniqueness of a solution to the system of equations (39) or the uniqueness of an optimum of the maximization problem (40).

Suppose we have found the marginal rate of substitution between each pair of final goods at every point in the n-dimensional final-goods space. The rest of the procedure for finding the utility function is familiar. Let

$$R_{i1}(z) \equiv R_i(z), \quad i = 2, 3, \dots, n$$

denote the marginal rate of substitution between the first and the ith final good at point z. Then the utility function can be obtained as a solution to the following total differential equation

$$dz_1 + R_2(z)dz_2 + R_3(z)dz_3 + \dots + R_n(z)dz_n = 0$$

This equation has a solution

$$U = U(z_1, z_2, \dots, z_n) \tag{44}$$

which is unique up to monotone transformation, if the $R_i(z)$ are continuously differentiable and satisfy the integrability conditions

$$\frac{\partial R_i}{\partial z_1} R_{i'} - \frac{\partial R_{i'}}{\partial z_1} R_i + \frac{\partial R_i}{\partial z_{i'}} - \frac{\partial R_{i'}}{\partial z_i} = 0, \qquad i, i' = 2, \dots, n$$

These integrability conditions are known to be satisfied if the consumer is rational in the sense that he has transitive preference orderings.[5] Finally, by substituting the relevant household production functions in (44), we arrive at the utility function which includes public goods as well as private goods as its arguments:

$$V(x, y) \equiv U(f^1(x^1, y_1), f^2(x^2; y_2), \dots, f^n(x^n; y_n)) \tag{45}$$

In concluding, we have to note the limitations of the present approach. First, the fundamental assumption that there exists and is known a household production technology is quite strong. Second, we have implicitly assumed that the consumer's market behavior honestly reflects his preference orderings. This may not be true if the consumer is aware that the information he conveys through his market behavior will be used in supplying and financing public goods. In this case his market behavior will show a change between the period when he is not aware of the possibility of the information being used and the period when he is. Thus there will arise an additional task of picking the right data. In spite of these limitations, it is hoped that the present approach will constitute a step toward putting the theory of optimal public expenditure on a more operational basis.

---

[5] See Hendrik Houthakker and Samuelson (1950).

REFERENCES

D. Gale and H. Nikaido, "The Jacobian Matrix and Global Univalence of Mappings," *Mathematische Annalen*, 1965, *159*, 81–93.

H. S. Houthakker, "Revealed Preference and the Utility Function," *Economica*, May 1950, *17*, 159–74.

M. Kurz, "Experimental Approach to the Determination of the Demand for Public Goods," *J. Publ. Econ.*, Nov. 1974, *3*, 329–48.

K. J. Lancaster, "A New Approach to Consumer Theory," *J. Polit. Econ.*, Feb. 1966, *74*, 132–57.

R. A. Musgrave, *The Theory of Public Finance*, New York 1959.

R. F. Muth, "Household Production and Consumer Demand Functions," *Econometrica*, July 1966, *34*, 699–708.

H. Nikaido, "Convex Structures and Economic Theory," New York 1968.

I. F. Pearce, "A Further Note on Factor-Commodity Price Relationships," *Econ. J.*, Dec. 1959, *69*, 725–32.

———, "More about Factor Price Equalization," *Int. Econ. Rev.*, Oct. 1967, *8*, 255–70.

P. A. Samuelson, "The Problem of Integrability in Utility Theory," *Economica*, Nov., 1950, *17*, 355–85.

———, "The Pure Theory of Public Expenditure," *Rev. Econ. Statist.*, Nov. 1954, *36*, 387–89.

———, "Summary on Factor-Price Equalization," *Int. Econ. Rev.*, Oct. 1967, *8*, 286–95.

A. Sandmo, "Public Goods and the Technology of Consumption," *Rev. Econ. Stud.*, Oct. 1973, *40*, 517–28.

H. Shibata, "A Bargaining Model of the Pure Theory of Public Expenditure," *J. Polit. Econ.*, Jan./Feb. 1971, *79*, 1–14.

R. H. Strotz, "The Empirical Implications of a Utility Tree," *Econometrica*, Apr. 1957, *25*, 269–80.

Market Equilibrium Versus Optimum in a Model with Congestion: Note

By Oded Hochman*

In their paper "Market Choice and Optimum City Size," Edwin Mills and David de Ferranti constructed a model of a city with congestion costs and derived the optimal land use pattern for it. However, they did not investigate the market equilibrium solution of their model and compare it to their optimal solution, as was done by Yitzhak Oron, David Pines, and Eytan Sheshinski.

In this paper the market equilibrium solution of the model is obtained and compared with the optimum solution derived by Mills and de Ferranti. It is shown that when congestion tolls are fully paid, market equilibrium is equivalent to optimality, as is the case in the paper by Oron et al. When no congestion tolls are paid, the city spreads further out than in the optimal case. The radius of the city increases, and land allocated to transportation at each location increases as well. Oron et al. were not able to obtain this result in their model.

Apart from the land use pattern, the average total cost of the urban economic activity in the model (housing and transportation) is investigated as a function of the population size. It is proved that average total cost and marginal total cost increase with population in the optimal case for any given population size. This result disproves the statement by Mills and de Ferranti that an optimum city size exists.

I. The Equilibrium Model

Mills and de Ferranti describe an optimization model of a town with congestion. Using their assumptions and notations, let u designate the distance from the center of the town. Then $L_1(u)$ is the land allocated to residential use at distance u, $L_2(u)$ the land allocated for transportation at distance u, and $T(u)$ the number of commuters crossing a circle of radius u. Let $p(u)$ be the cost per mile per commuter at distance u. Due to congestion this cost increases with the number of commuters $T(u)$ at distance u, and decreases with the land allocated for transportation, $L_2(u)$. We assume, following Mills and de Ferranti and William Vickrey, that it has the functional form as in equation (1).

$$p(u) = \bar{p} + \rho_1\left[\frac{T(u)}{L_2(u)}\right]^{\rho_2} \tag{1}$$

Let a_1 be the number of residents living on a unit of residential land. We assume a_1 to be constant. If $N(u)$ is the number of residents at distance u from the CBD, then

$$N(u) = a_1 L_1(u) \tag{2}$$

Let θ, $0 \leq \theta \leq 2\pi$ be the section of town, measured in radians, available for residential and transportation activities, then (3) follows immediately.

$$L_1(u) + L_2(u) = \theta u \tag{3}$$

$T(u)$, the number of commuters crossing a circle of radius u, is equal to the number of residents living outside this circle, hence

$$T(u) = \int_u^{\bar{u}} N(u)du \tag{4}$$

where $\bar{u}$ is the radius of the town.

The problem Mills and de Ferranti solved is the minimization of the total social cost TC associated with the residential and commuting activity, where TC is given in (5):

$$TC = \int_\epsilon^{\bar{u}} (p(u)T(u) + R_A\theta u)du \tag{5}$$

where ϵ, the radius of the CBD, and R_A, the agricultural land rent, are constants and exogenous given numbers.

---

* Department of economics and Center for Urban Studies, Tel-Aviv University.

The necessary condition for optimum is given by equation (6), (see (8) in Mills and de Ferranti):

$$
\frac{d\left[\dfrac{T(u)}{L_2(u)}\right]}{du} = -\frac{a_1}{\rho_2} - \frac{a_1\bar{p}}{\rho_1\rho_2(\rho_2+1)} \cdot \left[\frac{T(u)}{L_2(u)}\right]^{-\rho_2} \tag{6}
$$

We will now proceed to investigate the market equilibrium case. Let $R(u)$ be the equilibrium rent function and $C(u)$ the price an individual pays when he crosses a circle with radius u. The total housing and transportation cost of an individual located at distance u from the center is then:

$$
\int_0^u C(x)dx + a_1^{-1}R(u) \tag{7}
$$

where the first term expresses transportation cost and the second is the cost of the residential land. In this model no other costs are involved. This total cost of an individual should be the same everywhere under competitive equilibrium; otherwise, there will be excessive demand in some locations and no demand in others. Hence, by differentiating (7) with respect to u, we obtain

$$
C(u) + a_1^{-1}(dR(u)/du) = 0 \tag{8}
$$

A second condition is that land rents in each location be equal to the value of the marginal product of land in the transportation sector. The transportation product is the saving of costs rather than the production of positive income. The total value of transportation produced at location u is given by $(-T(u)p(u))$. The equality between the value of marginal product of land and rent is expressed in the following equation:

$$
R(u) = \partial(-T(u)p(u))/\partial L_2(u) \tag{9}
$$

This equation implies that transportation is produced efficiently even if its product is not necessarily sold in the market.

Substituting from (1) for $P(u)$ in the above equation and then differentiating, we get

$$
R(u) = \rho_1\rho_2(T(u)/L_2(u))^{\rho_2+1} \tag{10}
$$

We first assume that individuals pay full congestion tolls—i.e., each individual who passes through a circle with radius u pays an equal share of the rent of the land used for transportation at this radius. The total land used for transportation in radius u is $L_2(u)$, and the number of people crossing this circle is $T(u)$; hence, the equal share that each individual pays is $R(u)L_2(u)/T(u)$. Then $C(u)$ with full congestion tolls paid is:

$$
C_o(u) = p(u) + R(u)L_2^o(u)/T_o(u) \tag{11}
$$

where the index o indicates the case with congestion tolls. Substituting (11) into (8), and then substituting in it $R(u)$ from (10) and $dR(u)/du$ by differentiating (10), we get

$$
\frac{d\left[\dfrac{T_o(u)}{L_2^o(u)}\right]}{du} = -\frac{a_1}{\rho_2} - \frac{a_1\bar{p}}{\rho_1\rho_2(\rho_2+1)} \cdot \left[\frac{T_o(u)}{L_2^o(u)}\right]^{-\rho_2} \tag{12}
$$

But this is exactly equation (6) which is also equation (8) in Mills and de Ferranti. This means that the solution to the optimal problem solved by Mills and de Ferranti is also a solution to the case of market equilibrium with full congestion tolls.

Let us now assume that no congestion tolls are paid, i.e.,

$$
C_e(u) = p(u) \tag{13}
$$

where e indicates the case without congestion tolls. We still assume efficiency in transportation production; i.e., (9) and (10) are still valid. If we substitute (13) into (8) and substitute for $dR(u)/du$ from (10) and for $p(u)$ from (1), we get

$$
\frac{dy_e}{du} = -\frac{a_1}{\rho_2(\rho_2+1)} - \frac{a_1\bar{p}}{\rho_1\rho_2(\rho_2+1)}\, y_e^{-\rho_2} \tag{14}
$$

where:

$$y = T(u)/L_2(u) \tag{15}$$

and y_e is y in the case with no congestion tolls.

We will now try to compare characteristics of the town with congestion tolls to those of the town without congestion tolls. By setting $R(\bar{u}) = R_A$ in (10), we get:

$$y_o(\bar{u}_o) = y_e(\bar{u}_e) \tag{16}$$

From (12) and (14) we learn that the decline in $y_o(u)$ when u decreases is greater than in $y_e(u)$; i.e., y_o increases more rapidly with $(\bar{u}-x)$ than y_e. Hence, for any given distance from the boundary of the city x,

$$y_o(\bar{u} - x) > y_e(\bar{u} - x)$$

This means that at the distance x from the boundary of town,

$$T_o(\bar{u}_o - x)/L_2^o(\bar{u}_o - x) > T_e(\bar{u}_e - x)/L_2^e(\bar{u}_e - x)$$

This implies that at least one of the following relations must always be fulfilled.

$$T_e(\bar{u}_e - x) < T_o(\bar{u}_o - x) \tag{17a}$$

$$L_2^e(\bar{u}_e - x) > L_2^o(\bar{u}_o - x) \tag{17b}$$

If (17b) is fulfilled, it means that at any location with equal distance from the radius of town, more land is allocated for transportation in the town without congestion tolls (town e) than in the one with congestion tolls (town o). Less land will be allocated for dwelling at each such location in town e and therefore fewer residents will live there (than in town o). Equation (17a) is thus implied. If, on the other hand, we assume (17a), then between the town boundary and any given distance x from the boundary, fewer residents will live in town e than in town o. This implies that less land for dwelling and more land for transportation is allocated at this location in town e than in town o, which implies (17b). Hence, both relations in (17) must be fulfilled simultaneously. This implies that if both towns are of the same population size, town e must be more spread out than town o, i.e.,

$$\bar{u}_e > \bar{u}_o \tag{18}$$

The reason for this is that (17) implies that at any given distance x from $\bar{u}$, the density of population in town e is lower than in town o. For the two towns to be of the same size, ϵ must be further away from $\bar{u}_e$ than from $\bar{u}_o$ which implies (18).

The economic reasoning behind this argument is straightforward. Since land for transportation is a free product in the case without congestion tolls, it is abundantly used—so at every location there is too much land for transportation. Hence, the city is bigger than the optimal city size for the same population.

II. Average Total Cost

It is interesting to investigate the behavior of the total housing and transportation cost in this model as a function of population size. Mills and de Ferranti refer briefly to this question in the last section of their paper. They state that the average total cost for a low level of population size in the optimal town decreases when the population increases up to a certain minimum level of average total cost (ATC). A further increase in population size will cause ATC to increase. Hence, the city size in that minimum level of ATC is the optimum city size. This result has been questioned in the past,[1] since there is nothing in the model that permits economies of scale. It is proved in the following that ATC and marginal costs increase with population size at any level of population. It follows that no optimum city size exists.

Let us consider an optimal town with any given population N. We now reduce this population by an arbitrary quantity ΔN, $0<\Delta N<N$. The reduction will be done in a specific way and in two stages. The town at the final stage will be a feasible town, with total cost reduced more than proportionally to the reduction in population. It follows immediately that in an optimal town of the same population, the total cost will be even lower. Hence, marginal and average costs

---

[1] David Pines was the first to mention to me his doubts about this result. Mills himself also expressed, in a private communication, his doubts concerning this result.

increase monotonously at any given level of population.

Let us clarify the original optimal town with population of N by the subscript a. In the first stage, we reduce the population of town a by ΔN to get town b with a population $N-\Delta N$. The reduction is done by removing ΔN families living farthest from the center, and leaving the rest of the population intact. We then get a town with a new radius $\bar{u}_b$ and population $N-\Delta N$. Let us further reduce $L_2(u)$ at each $\epsilon<u<\bar{u}_b$ by the proportion $\Delta N/N$. The total cost of this new town, b, is less than a proportional reduction of the total cost of town a.

To prove it, let us consider the total cost function (5) and write it as follows:

$$TC = \int_{\epsilon}^{\bar{u}} p(u)T(u)du + SR_A \tag{19}$$

where S is the total area belonging to the town between ϵ and $\bar{u}$. The term $p(u)$ is given by (1), with $\bar{p}=0$.

First we prove that $S_b<S_a(1-\Delta N/N)$ where S_i — total area of town i and $S_i=L_1^i+L_2^i$ and where:

$$L_1^i = \int_{\epsilon}^{\bar{u}_i} L_1^i(u)du \tag{20a}$$

— total residential area in town i

$$L_2^i = \int_{\epsilon}^{\bar{u}_i} L_2^i(u)du \tag{20b}$$

— total land for transportation in town i

$i=a, b, c$, is an index of town

$$L_1^b = \left(1-\frac{\Delta N}{N}\right)L_1^a$$

since residential land is proportional to population.

$$L_2^b(U) = (1-\Delta N \mid N)L_2^a(u) \qquad \text{for } \epsilon < U < \bar{U}_b$$

Hence,

$$L_2^b < \left(1-\frac{\Delta N}{N}\right)L_2^a$$

since $L_2^a(u)$ was reduced proportionally for $\epsilon\leq u<\bar{u}_b$ and eliminated completely for $\bar{u}_b<u\leq\bar{u}_a$. It follows that

$$R_AS_b < \left(1-\frac{\Delta N}{N}\right)R_AS_a \tag{21}$$

From the definition of $T(u)$ in (4), it follows that

$$T_b(u) = T_a(u) - \Delta N$$

Putting $T_a(u)$ outside the brackets, we get

$$T_b(u) = T_a(u)\left(1-\frac{\Delta N}{T_a(u)}\right)$$

By substituting N for $T_a(u)$ in the denominator, we get, since $N\geq T_a(u)$

$$T_b(u) \leq T_a(u)\left(1-\frac{\Delta N}{N}\right). \tag{22}$$

and equality holds only when $T_a(u)$ equals N or zero.

From equation (1) with $\bar{p}=0$, we get

$$p_b(u) = \rho_1(T_b(u)/L_2^b(u))^{\rho_2}$$

Dividing the numerator and denominator by the factor $(1-\Delta N/N)$, and then substituting in the result $L_2^a(u)$ and $T_a(u)$, we get

$$\begin{aligned} p_b(u) &= \rho_1[(T_b(u)/(1-\Delta N/N) \\ &\qquad /(L_2^b(u)/(1-\Delta N/N))]^{\rho_2} \\ &\leq \rho_1(T_a(u)/L_2^a(u))^{\rho_2} \end{aligned}$$

hence,

$$p_b(u) \leq p_a(u)$$

From the above and (19), it follows:

$$T_b(u)p_b(u) \leq T_a(u)p_a(u)\left(1-\frac{\Delta N}{N}\right) \tag{23}$$

and equality holds only when $T_a(u)$ equals N or zero. Hence,

$$\int_{\epsilon}^{\bar{u}_b} p_b(u)T_b(u)du < \left(1-\frac{\Delta N}{N}\right)\int_{\epsilon}^{\bar{u}_a} p_a(u)T_a(u)du \tag{24}$$

By summing up (21) and (24) and substituting (19) into the result, we get

$$TC(b) < (1 - \Delta N/N)TC(a) \tag{25}$$

However, our proof is not yet complete, since town b does not fulfill all the requirements of a town in the model; at every $\epsilon < u < \bar{u}_b$ in b, there is a greater part not belonging to the town than the $(2\pi - \theta)$ allowed. We therefore change town b into town c by moving people from the town limits into the surplus vacant land inside the town. This transfer once again reduces TC by reducing $T(u)$ and $\bar{u}$, i.e., $T_c(u) < T_b(u)$ and $\bar{u}_c < \bar{u}_b$, while L_1 and $L_2(u)$ remain unchanged. Hence,

$$TC(c) < TC(b) < (1 - \Delta N/N)TC(a) \tag{26}$$

Town c fulfills the requirements of a town in the model. This completes the argument.

REFERENCES

E. S. Mills and D. M. de Ferranti, "Market Choice and Optimum City Size," *Amer. Econ. Rev. Proc.*, May 1971, *61*, 340–45.

Y. Oron, D. Pines, and E. Sheshinski, "Optimum Versus Equilibrium Land Use Patterns and Congestion Toll," *Bell J. Econ.*, Autumn 1973, *4*, 619–36.

W. Vickrey, "Pricing as a Tool in Coordination of Local Transportation," in *Transportation Economics*, Universities-Nat. Bur. Econ. Res. conference series, New York 1965.

Externalities, Extortion, and Efficiency

By GEORGE DALY AND J. FRED GIERTZ*

A number of major contributions to welfare economics in the last decade or so have emphasized the important role that can be played by voluntary exchange arrangements in achieving efficient levels of externality generating activities (see Ronald Coase; Ralph Turvey). These contributions seem to have been widely interpreted by their authors and others as questioning the necessity and/or desirability of collective intervention in many environments characterized by external effects. For example, in a paper summarizing this literature, Turvey writes:

> The first [main conclusion of the literature] is that if the party imposing the external diseconomies and the party suffering them are able and willing to negotiate to their mutual advantage, state intervention is unnecessary to secure optimal resource allocation. [p. 309]

More recently, however, such conclusions have been called into question. Critics have noted that the reliance on private bargaining to control the creation of externally harmful activities may well have the adverse effect of encouraging the very phenomena it seeks to control (see Jerome Rothenberg; E. J. Mishan). The possibility of the promotion of extortionate transactions casts doubt on the viability of bargaining as a means for the internalizing of externalities (see also Donald Shoup). In response to assertions of this sort, Harold Demsetz has recently argued that extortion constitutes a problem only if competition does not exist in the supply of extortion services. He has further stated, pp. 445–46, that extortion is a legal not economic concept, i.e., from an economic standpoint, it differs little from other types of transactions. It is apparent, then, that general agreement has not been achieved regarding either the likelihood of extortion or its probable impact on the allocation and distribution of resources in environments characterized by harmful external effects. Among the important issues which remain unresolved are the following: Does there exist a fundamental distinction between externality generating activities that do yield private benefits and those that do not, as Rothenberg, Mishan, and others argue? Or is such a distinction irrelevant from the viewpoint of economic analysis, as Demsetz suggests? If extortion is a meaningful category, does the existence of competition in its provision significantly affect its impact on allocative and distributive activities? The purpose of this paper is to examine these and related questions concerning the nature and consequence of the bargaining process when external effects are present.

I. External Harm and Bargaining: Some Neglected Aspects

Traditional (i.e., Pigovian) discussions of external effects have implicitly assumed that negotiation or bargaining is not possible between the parties involved; that "independent" behavior will characterize those parties involved in an externality relationship.[1] This assumption of independent behavior means that individuals are concerned only with their own choice variables and are ignorant of and/or indifferent towards the impact that behavior will have on others and, hence, will have no incentive to undertake activities which exert external harm but fail to yield private benefits. Thus, the assumption of independence precludes the undertaking of "bribe-motivated" activities be-

---

* Economics departments, University of Houston and Miami University (Ohio), respectively. The paper has benefited substantially from the comments of two anonymous referees.

[1] For a discussion of the nature and allocative significance of such "independent" behavior, see the related series of papers by James Buchanan and Milton Kafoglis, William Baumol, and Mancur Olson and Richard Zeckhauser.

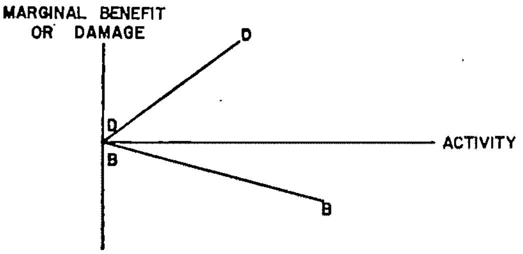


FIGURE 1

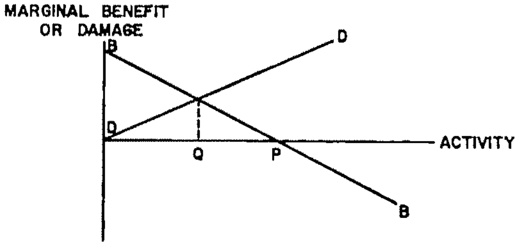


FIGURE 2

cause of the impediments to bargaining it implies.

It is this assumption of independent behavior which has been attacked in the more recent "Coase-inspired" literature. Yet this literature, while emphasizing the possibility of negotiations, has failed to note a vital fact: if bargaining *is* possible there is no a priori reason for ignoring the existence and consequences of externally harmful activities which fail to yield direct private benefits but can yield profitable side payments. That is, if bargaining is possible in cases where direct private benefits accrue it may also be possible where they do not. The economics of this latter case, ignored in the conventional literature, forms the main focus of this paper.

The term extortion refers in a general way to the act of obtaining payments from some entity in return for *not* imposing upon that entity some harmful effect, where the generator of the external effect receives no direct net internal benefit from the act. It is very important to distinguish between what we have labeled extortion and other bargaining situations where the externality generating party does receive direct net benefits.

Certain aspects of and contrasts between these cases can be clarified with the aid of some simple geometrical constructs. Consider a two-person world in which one of the individuals can engage in an activity which directly harms the other. Suppose that we measure the level of this activity on the horizontal axis and measure (in terms of some numeraire commodity) the evaluation that both individuals place upon this activity on the vertical. We abstract from income effects by assuming that for both individuals the marginal utility of the numeraire commodity is constant so that the individuals' marginal evaluation schedules can be used as demand curves.

In Figure 1, we measure the potential victim's marginal evaluation of the damage caused by the activity by schedule *DD* in the positive quadrant. Such a schedule shows the maximum amount this individual would pay to avoid the damage caused by an additional unit of the activity. The schedule *BB*, on the other hand, measures the direct marginal benefits net of cost to the individual whose activity imposes the external harm. It is in the negative quadrant indicating that the externality generating activity is costly and yields no direct benefits. With the possibility of bargaining (i.e., nonindependent behavior), such a situation clearly presents the potential for extortion with one person, the "victim," paying the other, the "extortionist," to refrain from an activity which yields the extortionist negative direct benefits. Successful negotiations, i.e., extortion, will lead to zero units of the activity produced and in absence of transactions costs would achieve Pareto optimality.[2]

Figure 2, on the other hand, illustrates the standard bargaining situation discussed in the externalities literature. In the absence of any impediments to bargaining, negotiations would lead to point *Q* where, as in Figure 1, joint benefits would be maximized and Pareto optimality would result. It is useful to note that in the context of Figure 2, bargaining has the potential for improving the allocation of resources as compared with the situation resulting from independent behavior. In contrast, extortionate bargaining (Figure 1) at best has the potential for duplicating the allocation of resources which would obtain under independent behavior.[3]

---

[2] Under these circumstances, i.e., absence of transactions costs, the act of extortion simply redistributes a consumer's and/or producer's surplus but does not affect the allocation of resources. It should be noted that for this condition to hold there must be no loss to the victim other than the resources he gives up to the extortionist. If, of course, the threat itself directly lowers the victim's utility (independently of the payment he must make), then the extortion is not costless.

II. Extortion and Allocative Efficiency

Extortion is clearly a pejorative term. It is important to analyze why this type of mutually agreeable exchange is held by society to be undesirable while other types of bargaining are relied upon in many circumstances to perform a socially useful task of producing a more efficient allocation of society's resources. While extortion redistributes wealth and is sometimes condemned on the basis of equity, there also seem to be important allocative reasons for its disfavor as well which relate to the costliness of the bargaining process.

Any bargaining process is a resource-using process.[4] Thus, in evaluating its effects on social welfare, the potential allocative gains resulting from that process must be weighed against the transactions costs generated. In each instance we must ask, "What are the allocative gains produced?" and "What are the costs incurred in producing them?" As Coase has pointed out, the Pigovian presumption that negotiations will not materialize in externality situations is tantamount to assuming that in each instance the relevant transactions cost will outweigh any potential allocative gains. It is also, as we have seen, to deny one of the necessary preconditions of extortion.

Consider these remarks in the context of Figures 1 and 2 above. As we have noted, the traditional Pigovian analysis has assumed "independent" behavior. This means that the individual will undertake each activity up to that point at which its net private benefits are zero. In Figure 1, independent behavior which implies the absence of bargaining would lead to zero units of the activity, while in Figure 2 it would result in P units. Thus, independent behavior produces inefficiency in the traditionally analyzed context of Figure 2 while, on the other hand, it produces efficiency in the case presented in Figure 1. In other words, *it is precisely the characteristic of independent behavior which produces nonoptimal allocation in the traditional externality relationship that produces optimality in situations where net private benefits do not exist.* Ironically, situations involving publicness and the attendant free-rider problems which often preclude the achievement of optimality through bargaining in traditional analyses may, in this situation, guarantee optimality by precluding effective negotiations and thus extortion.

Next, consider the social usefulness of the bargaining process in the two situations discussed above. As we have already noted, this process can lead to potential improvements in social welfare when the allocative gains it provides exceed the transactions costs it imposes. Those allocative gains, of course, are measured with reference to the allocation of resources in the absence of bargaining, that is, with independent behavior.

Viewed in this light, a fundamental distinction between the cases portrayed in Figures 1 and 2 emerges. Conventional bargaining (Figure 2) can lead to a superior resource allocation and therefore, depending upon the size of the transactions costs, may improve social welfare according to the Pareto criteria. On the other hand, the possibility of bargaining (i.e., extortion) in the context of Figure 1 can under the most favorable conditions (with zero transaction costs) only duplicate the results of independent behavior which produces allocative efficiency directly. With any positive level of bargaining costs, extortion will clearly lead to a reduction of social welfare since scarce resources are utilized in the process of negotiation while failing to improve the allocation of resources.

---

[3] The existence of extortion as defined above presupposes the satisfaction of several preconditions in addition to the existence of the potential extortionist and victim(s). The first, as we have noted, is that bargaining of some sort be possible between the parties involved. Secondly, it requires that the potential victim(s) enjoy a surplus from whatever activity the extortionist threatens. If such a surplus did not exist, the potential victim would be unwilling to pay the extortionist any positive sum for his services. Finally, given the existence of some extortable surplus and in a world of perfect foresight, the possibility of successful extortion also requires the existence of a technology and set of factor prices such that the costs to the extortionist of producing a credible threat are less than the "extortable" surplus. If this were not the case, extortion, although technically possible, would be unprofitable and thus in a world of rationality would not be undertaken.

[4] Several authors have noted that bargaining with respect to external effects is essentially a nonconstant sum game (see especially, Shoup). Since such games do not have determinant solutions, it is sometimes concluded that the associated transactions costs are infinite. Such a conclusion is clearly dubious and is not used in what follows. All that is assumed is that transactions costs are greater than zero.

It would appear then that there is a fundamental asymmetry between the social benefits of bargaining in the two cases, an asymmetry overlooked by economists' persistent tendency to exclude the case we have called extortion from their consideration. This tendency is particularly surprising since, as we have noted, situations involving external diseconomies and in which normal bargaining is possible are also likely to involve at some activity levels the possibility of extortion as well (for example, beyond point *P* in Figure 2). To the extent that this is so, the economists' case for negotiations in externality situations is of course correspondingly weakened.[5] Further, it is paradoxical that those authors who have stressed the possibility of negotiations have ignored extortion while, on the other hand, those who seem most skeptical of the possibility of negotiations in externality contexts have stressed its likelihood.

Finally, the above consideration suggests that in contrast to Demsetz's assertions, the distinction between extortion and other forms of bargaining has an economic as well as legal basis. Legal definitions of extortion seem to correspond rather closely to our Figure 1 situation (see Glen Mumey).[6] It would thus appear that the legal definition of extortion has a rather firm economic basis which in fact distinguishes it from other forms of bargaining. That is, the legal term extortion is applied to bargaining which can, like other forms of bargaining, redistribute wealth but which, unlike these forms, cannot improve the allocation of resources.

III. Extortion, Competition, and Efficiency

What effect will competition have upon the supply of extortion? Will it drive it out of existence as Demsetz (p. 445) seems to suggest? In more general terms, will it minimize the role played by extortion in the allocation and distribution of resources? Surely the issue is very complex and a thorough discussion of it would require an investigation of many complex, game theoretic consequences which might result from any negotiations. Nonetheless, several useful generalizations can be made.

First, as demonstrated in the preceding section, the effect of extortion on allocative efficiency depends upon the existence of transactions costs. With zero transactions costs, extortion or any other form of voluntary exchange simply redistributes wealth but does not use resources. In such a world all forms of voluntary bargaining are equally efficient. However, in a world of costly transactions, this is not so. Here extortion of the sort discussed in this paper uses resources and redistributes wealth but fails to improve allocative efficiency.

Of course, the simple fact that extortion involves costs does not necessarily imply that it is an undesirable method of allocating resources (see Gordon Tullock). As Coase has noted, all allocative methods involve costs. To accept or reject extortion on allocative grounds, we must know the nature of these alternative methods and, specifically, of their costliness. For example, extortion might well be the most efficient technique if the definition and enforcement of property rights to the amenities in question through the legal system and the prohibition of extortion were more costly than the process of extortion itself. On the other hand, if extortion could be prevented at zero cost, it would clearly be an inefficient means of allocating resources as noted above.

---

[5] This is not to argue, however, that collective intervention will necessarily improve things.

[6] Further, studies of actual cases seem to bear out such an assertion. For example, in a paper dealing with the question of conspiracy, R. A. Eastwood summarized the legal precedents as follows: "On this there has been much litigation, which I have not time to discuss in detail, but the results of the decisions, in my opinion, comes to this: that where the predominant or main purpose of the combination is not that of injury to others but the promotion of legitimate trade interests of those combining there is no tort, although others may incidentally suffer" (pp. 100–11).

What is clear, however, is that when extortion is a resource-using process the existence of competition in its supply is unlikely to either minimize its role or enhance allocative efficiency. Demsetz has argued that competition will drive to zero the price of extortionist services and, hence, that it will not be supplied. This confuses cause and effect. It is precisely the fact that extortionist services *are* supplied that drives the price to zero. Indeed, to argue as Demsetz has is to imply that a competitive industry in equilibrium will supply nothing since profits are zero. Moreover, if there are positive costs for the extortionist in establishing a credible threat, the price under competition will now approach these costs and not zero. In this circumstance the competitive provision of extortion services will clearly involve a net loss in social welfare by maximizing the resources devoted to this task. This would appear to lead to the worst of all possible worlds.[7]

IV. Conclusion

In a world of costless transacting, extortion (or any other form of bargaining) will not inhibit efficient resource allocation under either competitive or monopolistic circumstances; it may, however, result in profound changes in the distribution of income. When the process of extorting is a resource-using process, its existence will influence the allocation of resources as well as the distribution of income. Moreover, under these circumstances it is likely that competition will increase the use of resources and make extortion an extremely expensive method of allocating and distributing resources.

REFERENCES

W. Baumol, "External Economies and Second-Order Optimality Conditions," *Amer. Econ. Rev.*, June 1964, *54*, 358–72.

J. Buchanan and M. Kafoglis, "A Note on Public Goods Supply," *Amer. Econ. Rev.*, June 1963, *53*, 403–14.

R. Coase, "The Problem of Social Cost," *J. Law Econ.*, Oct. 1960, *3*, 1–44.

H. Demsetz, "Theoretical Efficiency in Pollution Control: Comment on Comments," *Western Econ. J.*, Dec. 1971, *9*, 444–46.

R. A. Eastwood, "Trade Protection and Monopoly," in G. W. Keeton and G. Schwarzenberger, eds., *Current Legal Problems*, Vol. 3, London 1950.

E. J. Mishan, "The Postwar Literature on Externalities: An Interpretive Essay," *J. Econ. Lit.*, Mar. 1971, *9*, 1–28.

G. A. Mumey, "The Coase Theorem: A Reexamination," *Quart. J. Econ.*, Nov. 1971, *85*, 718–23.

M. Olson and R. Zeckhauser, "The Efficient Production of External Economics," *Amer. Econ. Rev.*, June 1970, *60*, 512–17.

J. Rothenberg, "The Economics of Congestion and Pollution: An Integrated View," *Amer. Econ. Rev. Proc.*, May 1970, *60*, 114–21.

D. C. Shoup, "Theoretical Efficiency in Pollution Control: Comment," *Western Econ. J.*, Sept. 1971, *9*, 310–13.

G. Tullock, "Welfare Costs of Tariffs, Monopolies, and Theft," *Western Econ. J.*, June 1967, *4*, 224–32.

R. Turvey, "On Divergencies Between Social Cost and Private Cost," *Economica*, Aug. 1963, *30*, 309–13.

---

[7] In this context, it is instructive to note that organized crime's involvement in extortion often provides protection not just from their own imposition of the threatened action but from those of others as well. The elimination of competition in extortion does in fact provide a useful service in this case.

A Model of IQ, Occupation, and Earnings

By WILLIAM W. BROWN AND MORGAN O. REYNOLDS*

The objective of this paper is to present a simple model of causal relationships between three variables—IQ, occupation, and earnings. Relevant factors such as opportunity, risk attitudes, human investment, and motivation do not appear in the analysis, but the model produces implications which appear to be partially confirmed in preliminary evidence. Specifically, we present a model of the dependence of earnings upon IQ within and between occupations, show its implications for groups with different IQ distributions, and offer preliminary empirical tests with data from previously published studies.

I. The Model

Let income for the ith individual in the jth occupation be determined according to

$$Y_{ij} = a_j + b_j(IQ_i - t_j) + e_{ij} \tag{1}$$

where t_j is the required minimum or threshold IQ in occupation j; a_j is a parameter equal to average income when IQ is t_j in occupation j; b_j is a parameter>0; and e_{ij} is a random disturbance uncorrelated with IQ, with an expected value of zero, and a constant variance equal to $\sigma^2_{e_j}$.

Let a hierarchy of n occupations be defined by

(i) $t_1 < t_2 < \ldots < t_n$

(ii) $b_1 < b_2 < \ldots < b_n$

(iii) $a_1 < [a_1 + b_1(t_2 - t_1)] < a_2$
$< [a_2 + b_2(t_3 - t_2)] < a_3 < \ldots$
$< [a_{n-1} + b_{n-1}(t_n - t_{n-1})] < a_n$

That is, IQ thresholds and income coefficients increase with occupation. The last condition—that average income at threshold IQs exceed average income for those with identical IQs in the next lowest occupation—rules out intersections of average earnings functions for different occupations.

Now assume a normal distribution of IQ in the population and random assignment to occupations. That is, individuals are randomly drawn from the population and are randomly assigned to an occupation (all occupational assignments are equiprobable). If an individual's IQ is below the relevant threshold, he is not admitted to the occupation and is returned to the population for possible reassignment. If a randomly drawn individual's IQ is equal to or greater than the threshold of his randomly assigned occupation, he is employed in that occupation and his earnings are determined by (1). Under these conditions we have the following implications:

1. Average IQ in higher occupations will exceed average IQ in lower occupations.
2. Variation in IQ, measured by range or variance, will be greater in lower occupations.
3. Average earnings will be greater in higher occupations.
4. Implications 1 and 2 imply a negative correlation between mean IQ and IQ variation across occupations.

The first implication is simply an effect of truncation, where the conditional expectation of IQ monotonically increases with the value of the threshold t.[1] The second implication is obvious with respect to the range of IQs, but not for the other common measure of dispersion, variance. Following Arthur Goldberger we can show numerically

---

* Assistant professor of economics, California State University, Northridge, and Texas A&M University, respectively. We wish to acknowledge the help of Dennis Aigner, John Conlisk, Arthur Goldberger, Eugene Smolensky, an anonymous referee, and the managing editor. They are not responsible for any errors remaining in the paper. An earlier version was presented at the meetings of the Western Economic Association, August 1973.

[1] A technical appendix, available upon request, illustrates implications 1–9.

that the variance in a truncated normal distribution declines monotonically with increasing truncation points up to +5.0 standard deviations beyond the original first moment.[2] Implication 3 is a statement about the conditional expectation of earnings and follows from the assumed earnings functions.

In addition to implications 1–4, one additional proposition may also be true, given that certain empirical magnitudes exceed others. Specifically,

5. Earnings variation will generally be greater in the higher occupation.

Proposition 5 can be demonstrated by stating the variance of earnings in the *j*th occupation as

$$\mathrm{Var}\,(Y_j) = b_j^2\,\mathrm{Var}\,(\mathrm{IQ}_j) + \sigma^2_{e_j}$$

If we assume that error variances are equal across occupations, proposition 5 requires that the steeper rewards (b_j) in higher occupations offset smaller IQ variances.

II. Implications for Groups with Different IQ Distributions

Consider two subpopulations, *A* and *B*, whose IQ distributions differ; specifically, normal distributions in which the mean IQ of *A* is less than that of *B* and the variance for *A* is equal to that for *B*. Given the random assignment mechanism and occupational hierarchy specified in Section I, we have the following implications:

6. If the population ratio of *A* to *B* is *r*, then the ratio of *A* in all occupations will be less than *r*. Also, the relative representation of *A* will progressively decline with higher occupations.

7. Assuming that all individuals with the same IQ earn equal incomes on the average (i.e., discrimination against either group does not reduce incomes), the average earnings of *A* will be less than *B* within all occupations.

8. Within an occupation *A* will have smaller variability measured by range or variance in both IQ and income.

9. Within an occupation the correlation between earnings and IQ will be less for *A* than for *B*.

10. Without controlling for occupation, the regression coefficient of earnings on IQ will be lower for *A* than for *B*.

Implication 6 is simply a property of the normal distribution in which the effects of group differences in the mean are magnified in the relative proportions in the tails of the respective distributions. All *A*s and *B*s are not employed because the IQ threshold for the lowest occupation necessarily excludes some people from employment. An IQ of, say, 65 may be required in the lowest occupation, which would exclude relatively larger numbers of *A* from the labor force.[3] That is, below this IQ *A*s exceed their population share and, by arithmetic necessity, above it *A*s are below their population share. Similarly, the higher an IQ threshold for an occupation, the lower will be the relative proportion of *A*s who are eligible. Implication 7 follows from the fact that within any occupation, the conditional expectation of IQ will be lower for *A* than for *B*. Implication 8 follows because for any truncation point a larger proportion of the *B* population is eligible.

We can demonstrate implication 9 as follows. In general, for a group *g* in occupation *j*, the correlation between income and IQ, $\rho(Y_{j,g}, \mathrm{IQ}_{j,g})$, is given by

$$b_j\sigma_{\mathrm{IQ}_{j,g}}/\sqrt{b_j^2\sigma^2_{\mathrm{IQ}_{j,g}} + \sigma_e^2}$$

Since $\sigma^2_{\mathrm{IQ}_{j,B}} > \sigma^2_{IQ_{j,A}}$, $(1+\sigma_e^2/b_j^2\sigma^2_{IQ_{j,B}})$ is less than $(1+\sigma_e^2/b_j^2\sigma_{IQ_j,A}^2)$. Rewriting the last two terms as fractions, taking reciprocals and square roots proves that the correlation for *B* will exceed that for *A*. Implication 10 follows because sampling from *A* and *B* would yield proportionately more *A*s in lower occupations where the effect of IQ upon earnings is relatively small, and regressing earnings on IQ yields lower IQ coefficients in the absence of controlling for occupation.[4]

---

[2] Goldberger shows the decline in conditional variances up to +2.0 standard deviations. Tables for the standard normal probability function allow further calculations to +5.0 standard deviations. The conditional variance for the standard normal distribution at +2.0 standard deviations is .11 and at +5.0 standard deviations it is .04. Conditional variances probably continue to decline beyond that point but we have been unable to prove it analytically.

[3] In addition, all *A*s and *B*s are not employed because the random assignment mechanism specified in the model is a sampling procedure, not one that operates until complete exhaustion of the population. This is a legitimate specification on its own terms but also note that subsequent data are samples rather than universes.

III. Some Evidence

Among implications 1 to 4, perhaps the most interesting is the prediction that mean IQ and IQ variation will be negatively correlated across occupations. One set of data, from Cyril Burt, consists of a large representative sample of adults in England. The occupational classification was based upon the "degree of ability required for the work," not upon prestige or income. The simple correlation between mean IQ and standard deviation across the six occupational classes is −.93. Another data set from Thomas Harrell and Margaret Harrell consists of the scores on the Army General Classification Test (*GCT*) for 18,782 white enlisted men in the Army Air Forces Air Service Command during World War II. The range of *GCT* is generally narrower in higher *GCT* occupations. For example, the highest occupation is accountant with a mean score of 128.1 and a range from 94 to 157 while the lowest occupation, teamster, has a mean of 87.7 and a range of 46 to 145. The correlation between mean score and standard deviation is −.89. World War II data from Naomi Stewart also confirm the fourth implication. The Army *GCT* scores for 81,553 white enlisted men are arrayed across 227 different occupations and the correlation between the median score and the semi-interquartile range across occupations is −.80. Leona Tyler (pp. 336–40) cites other data which confirm this prediction.[5]

Data from Phillips Cutright offer a source for preliminary testing of predictions about group differences in IQ and earnings. The sample consists of 25,000 white and 4,000 black men with Armed Forces Qualification Test (*AFQT*) scores during the Korean War. Subsequent Social Security data were matched to obtain earnings and educational attainment up to 1964. Table 1 presents some of the main features of these data. Because these data do not have occupational classifications, we can interpret education measured by years as a crude proxy for occupation.[6] Table 1 shows that the black mean *AFQT* and educational attainment are nearly one standard deviation below the respective white averages while the black earnings mean is about two-thirds of a standard deviation below that for whites. White standard deviations are substantially larger for *AFQT* and earnings, while the standard deviation for education is marginally higher for blacks. All correlations between variables are higher for whites than blacks, except those between earnings in different years. The correlation between *AFQT* and education is about 50 percent higher for whites than for blacks. The correlations between *AFQT* and earnings are

---

[4] While the model implies that the earnings-IQ slope will be higher for *B* than *A* independent of occupation (implication 10), whether the simple correlation between earnings and IQ will also be higher for *B* is an empirical question. Recall that the slope equals $\rho\sigma_Y/\sigma_X$, where σ_Y is the standard deviation of earnings and σ_X is the standard deviation of IQ. Alternatively, we can write $\rho=b\sigma_X/\sigma_Y$. The numerator is larger for *B* than for *A* but the denominator is also larger for *B*. Subsequent data show a substantially higher correlation between earnings and IQ for *B* (whites) than for *A* (blacks).

[5] The variance of IQ scores within occupations implies that correlations between individual occupational rank and IQ will be far from perfect. For example, these correlations are only .42 and .45 in the Harrell and Stewart samples, respectively. But because these samples were mostly for young men, the correlations are likely to be lower than in older civilian populations. The older a cohort is, the more likely it is that high IQ individuals have gravitated toward higher occupations and low IQ individuals toward (remain in) lower occupations. Data from the U.S. Employment Service (see Otis Duncan, David Featherman, and Beverly Duncan, pp. 98–101) for a civilian population has a correlation of .55 between intelligence and occupational status, about the same as the correlation between intelligence and scholastic achievement. Tyler, p. 343, also cites studies that IQs of school boys correlated .57 and .71 with their occupational status 14 and 19 years later, respectively.

[6] Educational attainment is generally correlated with occupation about 0.6.

TABLE 1

Variable	Mean	Standard deviation	Correlation Coefficients 1	2	3	4	5	6
Whites With 1962–64 Earnings ($N=24{,}819$)								
1. *AFQT*	52.41	30.75	1.000	.683	.236	.337	.350	.327
2. Education	10.66	2.68		1.000	.256	.335	.334	.322
3. 1958 Earnings	$3800	$2445			1.000	.596	.585	.544
4. 1962 Earnings	5378	3393				1.000	.827	.778
5. 1963 Earnings	5645	3605					1.000	.827
6. 1964 Earnings	6077	3964						1.000
Blacks With 1962–64 Earnings ($N=4008$)								
1. *AFQT*	14.10	16.95	1.000	.483	.083	.112	.096	.126
2. Education	8.45	2.80		1.000	.089	.128	.111	.148
3. 1958 Earnings	$2139	$1597			1.000	.660	.600	.566
4. 1962 Earnings	2948	2160				1.000	.863	.782
5. 1963 Earnings	3131	2238					1.000	.873
6. 1964 Earnings	3481	2447						1.000

Source: Cutright, p. 28.

about three times higher for whites (see fn. 4).

Table 2 presents a cross-classification of the Cutright data by education and *AFQT* score. The column totals illustrate the prevailing differences in *AFQT* distribution between whites and blacks. If we take the liberty of interpreting the five education rows as occupational groups, it is apparent that the relative representation is most different in the two extreme classes, with blacks three times more prominent in the lowest strata and whites four times more prominent in the top strata, as required by implication 6. Mean earnings generally increase with *AFQT* and education for both

TABLE 2—FREQUENCY DISTRIBUTION FOR MEN WITH 1962–64 EARNINGS, BY EDUCATION, *AFQT*, RACE, AND EARNINGS

Years of Education	*AFQT* 0–9	10–29	30–49	50–79	80–99	Total	1964 Mean Earnings
White, $N=25{,}029$							
0– 7	5.6	5.1	1.0	.3	.1	12.1	$3960
8– 9	2.7	7.5	4.3	2.5	.6	17.6	5035
10–11	.7	4.6	4.8	5.2	2.2	17.5	5597
12	.3	3.0	6.6	13.8	11.6	35.3	6371
13+	.1	.4	1.3	4.0	11.3	17.2	8479
Total	**9.4**	**20.7**	**18.0**	**26.1**	**25.8**	**100.0**	**$6068**
1964 Mean Earnings	$3847	$4893	$5700	$6248	$7862		
Black, $N=4{,}059$							
0– 7	28.0	7.3	0.5	.2	.2	36.4	$3085
8– 9	15.8	8.3	1.0	.2	.0	25.4	3381
10–11	8.8	10.0	2.0	.7	.0	21.7	3836
12	1.8	5.0	2.7	1.7	.3	11.6	3887
13+	0.0	1.2	1.2	1.3	.7	4.4	4647
Total	**54.5**	**32.3**	**7.5**	**4.2**	**1.3**	**100.0**	**$3489**
1964 Mean Earnings	$3304	$3616	$3510	$4185	$5617		

Note: Totals have been rounded out.
Source: Calculated from Cutright, p. 37.

groups as is generally expected, but within educational classes the average earnings of blacks are uniformly less than those of whites, which is consistent with implication 7. In addition, the association between *AFQT* and mean earnings is not as strong for blacks as for whites, which can easily be seen by inspection of Table 2.[7] The simple regression of 1964 earnings on *AFQT* yields a coefficient of $42 for whites and $18 for blacks, as predicted by implication 10. The coefficient is similar to the $25 obtained by W. Lee Hansen, Burton Weisbrod, and William Scanlon in a multiple regression of 1963 earnings upon *AFQT* and other variables for a national sample of 2400 low-achieving young men. Their sample was 47 percent white, 53 percent black, and had a mean *AFQT* of 9.1. Empirical work by John Hause (1971, 1972) is also broadly consistent with our simple IQ model. Hause uses four sets of data in regression analyses and argues that earnings coefficients for mental ability increase with education level, that is, IQ has very weak associations with earnings at lower educational levels. He also argues that IQ coefficients tend to become more important as labor force experience increases.

IV. Conclusion

The implications of our simple model require much greater testing against the Cutright data and other sources before preliminary acceptance is warranted, but the evidence does not contradict the model in any obvious way, despite its simple formulation. Some implications of the model might be interpreted as competing explanations for phenomena usually explained by various discrimination models. This is correct because the model helps to explain the occupational pattern for Chinese-Americans and Jews, as well as blacks, all of whom have been victims of discrimination but who occupy very different places in the occupational hierarchy. However, our model does not necessarily conflict with models which explain the labor market position of different ethnic groups by "discrimination." For instance, controlling for education and *AFQT* in the Cutright data eliminates less than half of the black-white earnings gap, which means that other differences exist between groups, perhaps discrimination and/or black disadvantages in other productivity characteristics.

To further illustrate that the labor market position of various groups cannot be entirely explained by IQ distributions, males and females have approximately equal IQ distributions but very different occupation-earnings distributions. Another example is that American Indians average about one-half of a standard deviation higher than the average for blacks on ability and achievement tests, but the occupational distribution is substantially lower for American Indians.[8] Nevertheless, a striking feature of the model is that substantial group differences in occupation and earnings occur if group differences in IQ distributions are coupled with a credible specification of the relationships between occupational earnings and IQ, even though a "discrimination-free" assignment mechanism is posited.

---

[7] Cutright does not report the correlations between *AFQT* and earnings within educational classes, which would permit a test of implication 9. But given that the correlations between *AFQT* and education, earnings and education, and earnings and *AFQT* are all substantially higher for whites than for blacks, it is improbable that the earnings—*AFQT* correlation within an educational class would be lower for whites.

[8] A related issue concerns the emerging evidence that there are stable ethnic and sex differences in patterns of mental ability (verbal, reasoning, number, space), which are invariant across wide socioeconomic classes (see Gerald Lesser, Gordon Fifer, and Donald Clark; Susan Stodolsky and Lesser; Tyler). The patterns of scores, in addition to a composite score like IQ, are bound to interact with occupational demands.

REFERENCES

C. Burt, "Intelligence and Social Mobility," *British J. Statist. Psych.*, May 1961, *14*, 3–24.

P. Cutright, *Achievement, Mobility, and the Draft: Their Impact on the Earnings of Men*, U.S. Social Security Admin. staff paper 14, Washington 1973.

O. D. Duncan, D. L. Featherman, and B. Duncan, *Socioeconomic Background and*

Occupational Achievement: Extensions of a Basic Model, final rept., proj. no. 5–0074 (E-191), U.S. Office of Education, Washington 1968.

A. S. Goldberger, "Selection Bias in Evaluating Treatment Effects: Some Formal Illustrations," Inst. Res. on Poverty disc. pap. series, Univ. Wisconsin, Apr. 1972.

W. L. Hansen, B. A. Weisbrod, and W. J. Scanlon, "Schooling and Earnings of Low Achievers," *Amer. Econ. Rev.,* June 1970, *60,* 409–18.

T. W. Harrell and M. S. Harrell, "Army General Classification Test Scores for Civilian Occupations," *Educ. and Psych. Measurement,* Autumn 1945, *5,* 229–40.

J. C. Hause, "Ability and Schooling as Determinants of Lifetime Earnings or If You're So Smart, Why Aren't You Rich?", *Amer. Econ. Rev. Proc.,* May 1971, *61,* 289–98.

———, "Earnings Profile: Ability and Schooling," *J. Polit. Econ.,* May/June 1972, suppl., *80,* S108–38.

G. S. Lesser, G. Fifer, and D. H. Clark, "Mental Abilities of Children from Different Social-Class and Cultural Groups," *Monogr. Soc. Res. Child. Develop. 30* (4), Chicago 1965.

N. Stewart, "A.G.C.T. Scores of Army Personnel Grouped by Occupation," *Occupations,* Oct. 1947, *26,* 5–41.

S. S. Stodolsky and G. S. Lesser, "Learning Patterns in the Disadvantaged," *Harvard Educ. Rev.,* Fall 1967, *37,* 546–93.

L. E. Tyler, *The Psychology of Human Differences,* 3d ed., New York 1965.

On Monopoly Welfare Losses: Comment

By R. CARSON*

In a recent article in this *Review*, Abram Bergson has challenged the thesis of Arnold Harberger, David Schwartzman, Michael Klass, Harvey Leibenstein, Dean Worcester, and others that the "welfare loss" from monopoly is apt to be inconsequential in a large economy, "such as that of the United States."[1] Instead, he argues that this loss may well be high, and that it "... is quite sensitive to empirical aspects which still seem unsettled ..." (p. 867). Bergson's contribution, which economists will applaud, is to measure the welfare loss within a general equilibrium framework. This is also the source of his challenge. Nevertheless, I shall argue that Bergson's conclusions stem largely from extreme assumptions about monopolistic pricing. When these assumptions are replaced by more "reasonable" ones, the small efficiency losses registered by previous writers are partially reconfirmed.

I. The Nature of the Monopoly Welfare Loss

Of course, firms in monopoly positions may do many things to cause welfare losses, broadly defined, and few economists would argue that the net impact of these together is necessarily small. However, in the conventional technical literature, the welfare loss from monopoly is solely an efficiency loss resulting from tendencies of imperfectly competitive firms (or cartel authorities) to set prices above marginal costs. Suppose that every good satisfies final consumer demands. Then the welfare loss from monopoly arises because the economy fails to produce the mix of goods and services that best satisfies consumer preferences. This is the loss measured by Harberger, Schwartzman, and Bergson. For a two-good economy, it is captured by Figure 1, basically a replica of Bergson's Figure 1. There the production-possibility frontier is assumed to be a straight line (analogous to constant long-run average costs), the community's preferences are described by indifference curves U_1 and U_2, and x_1 and x_2 index outputs in imperfectly and perfectly competitive sectors, respectively. Let P_1 and P_2 be the prices of these same goods.

If firms in both sectors set price equal to marginal cost, the economy will find its constrained bliss with the mix of goods denoted by A, and I_c will measure the community's income at competitive prices. On the other hand, if producers of x_1 set P_1 above MC_1

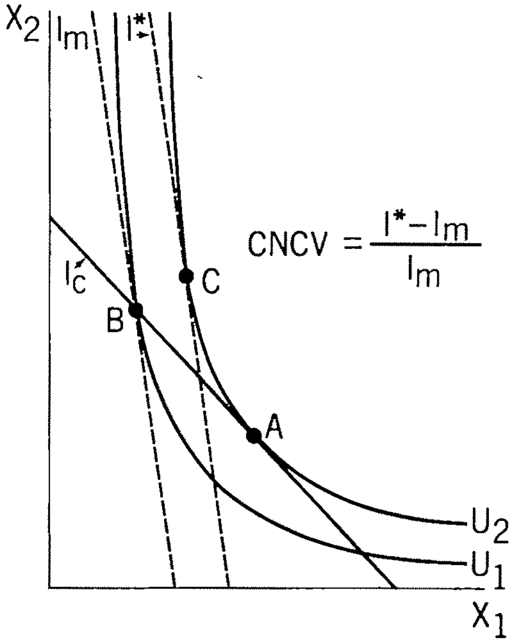


FIGURE 1

---

* Associate professor of economics, Carleton University. I am indebted to Jessie Lim for typing the first draft of the manuscript, to Josie Obina for typing the final copy, and to Thundercloud for drawing the diagram.

[1] See the references to Bergson, Harberger, Schwartzman, Worcester, Klass, and Leibenstein. Finally, see Frederic Scherer for an excellent summary discussion prior to Bergson. "Large" economy means specifically here that all industries and firms operate under constant costs.

while producers of x_2 set $P_2=MC_2$, the economy moves to the less-preferred combination at B. The income of the community at monopoly prices is given by I_m. If the level of welfare prevailing at A were to be restored at those prices, income would have to rise to I^*. As conventionally measured, the welfare loss from monopoly is the percentage difference between I^* and I_m or, in Bergson's terminology:

$$(1) \qquad CNCV = \frac{I^* - I_m}{I_m}$$

where $CNCV$ is the welfare loss in question.

The above is straightforward, but it requires an estimate of σ, the elasticity of substitution between the outputs of sectors one and two over segment AC of indifference curve U_2. Lacking this, empirical studies have used a partial equilibrium analysis instead, approximating the welfare loss as a loss in consumers' surplus, due to monopolistic pricing, not counterbalanced by increased profits in the imperfectly competitive sector. Bergson gives a lucid discussion of the error that may be built into such an approximation. I shall not repeat it here, but I show in the Appendix that for any *given* percentage markup of P_1 over MC_1 in the imperfectly competitive sector—or for any given production frontier and ratio of "with monopoly" prices—$CNCV$ is an increasing function of σ.[2] When all is said and done, this is the source of the high values of $CNCV$ that Bergson is able to generate.

II. Bergson's Model and Conclusions

One way to investigate the welfare loss from monopoly within a general equilibrium framework is to assume a community indifference map with the elasticity of substitution as a parameter, a particular production-possibility frontier, and particular price markups over marginal costs in monopoly sectors. Thus, Bergson supposes a linear production-possibility frontier, price markups that average 10 percent, 20 percent, and 30 percent, respectively, over MC, and an indifference map defined by the formula:

$$(2) \qquad U(x_1, x_2, \ldots, x_{n+1}) = \sum_{k=1}^{n+1} A_k x_k^{(1-1/\sigma)}; \qquad \text{for } \sigma \neq 1$$

Here, community welfare is some increasing function of U, the $\{A_k\}$ are constants, σ is the elasticity of substitution, assumed to be constant and the same between all pairs of goods, and the $\{x_k\}$ index production in different sectors of the economy. Perfect competition is assumed to hold in sector $(n+1)$ and imperfect competition elsewhere. Each firm is assumed to be vertically integrated from raw material supplies to final-use demand.[3] Let λ_k be the ratio of price to marginal cost in sector k and let γ_k be the percent of I_m (at monopoly prices) spent on x_k. Bergson also shows that:

$$(3) \qquad 1 + CNCV = \left(\sum_{k=1}^{n+1} \gamma_k \lambda_k^{\sigma-1}\right)^{1/\sigma-1} \left(\sum_{k=1}^{n+1} \frac{\gamma_k}{\lambda_k}\right)$$

for values of σ other than 1. He then proceeds to investigate $CNCV$ in two- and three-sector models, always assuming that the single competitive sector absorbs half of I_m.

In the two-sector case, Bergson calculated welfare losses (Table 1, p. 864) rise from .06 to 3.86 percent of I_m as σ increases from .5 to 64, with a monopoly price markup of 10 percent (i.e., $\lambda_1=1.1$). When $\lambda_1=1.2$, the welfare loss rises from .21 to 8.8 percent over the same range of σ, rising, say, from 1.66 to 3.20 as σ rises from 4 to 8. Finally with a monopoly markup of 30 percent, the welfare loss rises from .43 to 13.74 percent, reaching 1.73, 3.43, and 9.95 percent at σ values of 2, 4, and 16, respectively. To keep matters in perspective, it should be recalled that Harberger and

---

[2] Note that B in Figure 1 is the point where an indifference curve, whose slope is given by minus the ratio of the "with monopoly" prices, passes through the production frontier.

[3] If firms do not behave "as if" this were true, there may be additional welfare monopoly losses, due to price distortions in intermediate-goods industries, not captured in the discussion below. Klass has tried to measure these, using the *U.S.* 1958 input-output table. His results indicate that *at most* a 40 percent increase in welfare losses could have stemmed from this source.

Schwartzman calculated *CNCV* to be .1 percent of *GNP* or less, although a number of authors have claimed these to be underestimates.[4] In the two three-sector models considered by Bergson (Table 2, p. 865), *CNCV* is generally higher. For example, it can be as high as 11.54, 25.47, and 39.03 percent with $\sigma=64$ and price markups of 10, 20, and 39 percent, respectively, over marginal cost.

III. The Welfare Loss Recalculated

A problem with Bergson's approach is that firms are assumed to be completely oblivious to the elasticities of their demand curves. For any degree of competition in an imperfectly competitive industry, each firm will find that its demand elasticity rises with σ, a factor that would normally lead it to charge lower prices for higher values of σ. This could more than offset the above tendency for *CNCV* to rise with σ. Bergson is, in fact, often assuming that the ratios $(P-MC)/MC$, chosen by imperfectly competitive industries, will be many times those appropriate to profit-maximizing monopolies facing no threat of entry, government regulation, or countervailing power. In practice, the term monopoly must embrace a variety of forms of oligopoly, as well as the case where an industry is cartelized or consists of but a single firm. Even an absolute monopolist is likely to encounter government regulation or some other form of countervailing power that effectively limits its ability to mark up price above marginal cost.

In this context, let η be the elasticity of demand facing a particular industry, computed according to the formula:

$$\eta=\sigma-\gamma(\sigma-1) \tag{4}$$

where γ is again the percent of I_m absorbed by this industry. Let ϵ be the demand elasticity relevant to a particular price-setting decision maker within the industry. As a profit maximizer, he is usually assumed to try to set his price equal to:

$$\frac{\epsilon}{\epsilon-1}MC \tag{5}$$

which is a markup of $100/(\epsilon-1)$ percent over *MC*. Evidently, $\epsilon\geq\eta$, provided the monopoly does not reach beyond the "industry" defined above. Otherwise, we lack a specific link between ϵ and η unless we assume with Bergson that every identifiable product constitutes a separate good produced by a separate industry. Then $\epsilon=\eta$, but now the elasticity of substitution between the services of a Gremlin and a Pinto must be the same as that between the services of a Pinto and oranges.

At some cost of theoretical purity, we could instead take η to be the elasticity of demand for a product group and ϵ to be the (often higher) elasticity facing a "representative" producer within the group whose P/MC is average for the group as a whole. Bergson does not do this because he feels that such "horizontal aggregation" will conceal welfare losses. This conclusion follows from the higher losses recorded in his three-sector over his two-sector model, for given σ and average price markups in monopoly sectors (pp. 864–65). I shall return to this matter below.

Bergson also argues that (5) " . . . can be [no] more than very broadly applicable to real world monopolies . . . " (p. 861). This point is in fact crucial to his argument because accepting (5) as an upper bound to monopoly prices does not permit them to be raised indefinitely without regard for demand elasticity. Evidently, firms intending to maximize profits will sometimes err, setting prices too high or low on that account. But there is no reason for the expected value of these deviations to be anything other than zero. Moreover the standard error of this estimate in an economy with hundreds of thousands of decision makers setting prices and billions of goods to be priced is likely to be minuscule because of the law of large numbers.

In addition, there are at least six other reasons often mentioned in the literature why $MC=MR$ may fail to hold. The first of these is collusion, including price leadership, which for simplicity I assume to take the form of joint profit maximization over more than one product. However, this will not cause (5) to be violated (with $\epsilon \geq \eta$) if η is the elasticity of demand for a class of goods at least as broad as the extent of the collusion. This is again a question of whether horizontal aggregation across substitute products conceals welfare losses.

---

[4] See the discussion by Scherer.

The other reasons for $MC=MR$ to be violated are goals other than long-run profit maximization, the threat of entry by new competitors (or of government regulation), the durability of many goods, misleading advertising, and X-inefficiency. But Bergson's assumption that the economy operates along its production-possibility frontier rules out the last two, although the efficiency losses that they cause *may* be high in the real world. The remaining factors tend to depress P/MC, assuming that alternative management goals are more expansive than long-run profit maximization, for any given level of costs, as the thrust of the literature on the subject appears to suggest.[5]

In short, using (5) to define maximal price markups instead of assigning them arbitrarily as Bergson does should give us an outer bound for monopoly welfare losses, other than those stemming from X-inefficiency and misleading advertising—that is, as monopoly welfare losses are usually defined.[6] (This is subject to settling the problem of horizontal aggregation.) Refer to Tables 1–3. Table 1 describes a two-sector economy in which a competitive and a monopoly sector each absorb 50 percent of I_m. Tables 2 and 3 each describe three-sector economies in which the lone competitive sector continues to absorb 50 percent of I_m. In Table 2, each monopoly sector absorbs 25 percent of I_m. In Table 3, the first monopoly sector absorbs 33 percent and the second 17 percent of I_m. All of this is identical with Bergson's scenario. The formulas at the bottom of the tables, derived in the Appendix from formulas (3), (4), and (5) above, give $CNCV$ as a function of parameters m and p, which take on the values indicated.

TABLE 1—THE PERCENTAGE "WELFARE LOSS FROM MONOPOLY" ($CNCV$) IN A TWO-SECTOR ECONOMY[a]

σ	$m=1$	$m=2$	$m=3$
2	33.3	4.2	**1.6
4	13.0	**2.5	1.0
8	** 5.9	1.4	.6
16	2.8	.7	.3
32	1.4	.4	.2
64	.7	.2	.1

Note: σ = elasticity of substitution; η = elasticity of demand for monopoly sector; ϵ = elasticity of demand for "representative" product group within monopoly sector ($\epsilon = m\eta$).

[a] $$1+CNCV=\left[1-\frac{1}{m(\sigma+1)}\right]\cdot\left[\frac{1}{2}\left(1+\left(\frac{m(\sigma+1)}{m(\sigma+1)-2}\right)^{\sigma-1}\right)\right]^{1/\sigma-1}$$

The first column of each table assumes that monopoly prices do in fact equal $\eta/(\eta-1)$ times MC, and we note that these results reverse Bergson's in two critical respects. First, as σ rises, $CNCV$ *falls*. Second, as the number of sectors rises, $CNCV$ also falls. The importance of the latter lies in Bergson's claim that horizontal aggregation of products understates the the true monopoly loss. If anything, aggregation appears to *overstate* the true efficiency cost as a comparision of Table 1 with Tables 2 and 3 indicates (taking the first column in each case).

Because the evidence cited above does not indicate an exaggeration of "true" monopoly welfare losses due to horizontal aggregation, I shall consider each sector to be a product aggregate in the remaining columns of Tables 1, 2, and 3. That is, ϵ is the elasticity of demand facing a "representative" decision maker whose price-to-marginal-cost ratio for the good in question is average for the entire monopoly sector. Thus $\epsilon \geq \eta$ must hold and, having considered the case $\epsilon=\eta$, we now take ϵ to be a multiple of η. Thus, column 2 of Table 1 assumes $\epsilon=2\eta$, while column 3 assumes $\epsilon=3\eta$. In each case, $CNCV$ falls as σ rises, and, of course, $CNCV$ falls for given σ as ϵ becomes a larger multiple of

---

[5] For example, William Baumol's sales maximizers set $P=MC$ when constant costs prevail. See, as well, Armen Alchian and John Palmer.

[6] See William Comanor and Leibenstein for a discussion of welfare losses due to X-inefficiency. My own view is that they have greatly overstated these losses.

Table 2—The Percentage "Welfare Loss from Monopoly" (*CNCV*) in a Three-Sector Economy (with Equal Monopoly Weights)[a]

σ	$m, p=1$	$m=1; p=2$	$m=1; p=3$	$m, p=2$	$m=2; p=3$	$m, p=3$
2	19.0	12.6	12.7	2.8	2.1	1.1
4	6.7	4.9	4.9	1.4	**1.0	.6
8	2.9	**2.2	**2.2	.7	.5	.3
16	1.4	1.0	1.1	.3	.3	.2
32	.7	.5	.5	.2	.1	.1
64	.3	.2	.3	.1	.1	0

Note: σ = elasticity of substitution; η_k = elasticity of demand for monopoly sector $k(\eta_1=\eta_2=\eta)$; ϵ_k = elasticity of demand for "representative" product in monopoly sector $k(\epsilon_1=m\eta;\ \epsilon_2=p\eta)$.

[a] $$1+CNCV=\left[1-\left(\frac{m+p}{mp(3\sigma+1)}\right)\right]\left[\frac{1}{2}+\frac{1}{4}\left(\frac{m(3\sigma+1)}{m(3\sigma+1)-4}\right)^{\sigma-1}+\frac{1}{4}\left(\frac{p(3\sigma+1)}{p(3\sigma+1)-4}\right)^{\sigma-1}\right]^{1/\sigma-1}$$

η. Tables 2 and 3 take the same approach with $\epsilon_1=m\eta_1$ and $\epsilon_2=p\eta_2$. (In Table 2, $\eta_1=\eta_2=\eta$, because each monopoly sector has a weight (γ_k) of $\frac{1}{4}$. In Table 3 sector one has a weight of .33, while sector two has a weight of .17). Again, *CNCV* falls as σ rises, and the three-sector models have smaller monopoly welfare losses than the two-sector model in comparable situations. (As in Bergson's computations, increasing the dispersion of the sector weights raises *CNCV*.)

Because the low values of σ now yield high monopoly welfare losses, we must next ask how low σ can "reasonably" go within the context we are considering. For any given maximal markup of *P* over *MC* over all monopoly sectors, each column in each table gives us a value of η_k where k is the monopoly sector with the highest price markup. From (4), we can translate this into a lower bound for σ. If we insist on some variation in price markups between the two monopoly sectors in Table 2, we must add the constraint $m\neq p$. The same constraint is necessary in Table 3 if we insist on a variation in price markups of 25 percent or more between the

Table 3—The Percentage "Welfare Loss from Monopoly" (*CNCV*) in a Three-Sector Economy with Unequal Monopoly Weights[a]

σ	$m, p=1$	$m=1;$ $p=2$	$m=2;$ $p=1$	$m, p=2$	$m=1;$ $p=3$	$m=3;$ $p=1$	$m=3;$ $p=2$	$m=2;$ $p=3$	$m, p=3$
2	20.4	17.7	8.8	3.0	18.1	8.3	1.7	2.6	1.2
4	7.4	6.8	3.5	1.5	7.1	3.2	**.9	**1.4	.6
8	3.2	**3.1	**1.6	.8	**3.2	**1.5	.4	.7	.3
16	1.5	1.5	.8	.4	1.5	.7	.2	.3	.2
32	.7	.7	.4	.2	.8	.3	.1	.2	.1
64	.4	.4	.2	.1	.4	.2	.1	.1	0

Note: See note to Table 2.

[a] $$1+CNCV=\left[1-\frac{1}{(2.0303\sigma+1)m}-\frac{1}{(4.8824\sigma+1)p}\right]\left(\frac{1}{2}+.33\left(\frac{2.0303\sigma+1)m}{(2.0303\sigma+1)m-3.0303}\right)^{\sigma-1}+.17\left(\frac{p(.83\sigma+.17)}{p(.83\sigma+.17)-1}\right)^{\sigma-1}\right]^{1/\sigma-1}$$

two monopoly sectors, a variation that is necessary to approximate the degree of variability observed by Harberger.

I shall assume a maximal monopoly price markup of 20 percent, ($\epsilon_1 = 6$) clearly considered to be extreme (and rare) by Harberger and Schwartzman, authors of the most frequently cited empirical works on the subject of monopoly welfare losses. Let σ_{ji} be the lowest permissible value of σ in column j of Table 1. I have identified σ_{ji} by placing a ** in each column of the three-sector tables where $m \neq p$, as well as in the two-sector table.

As any developed economy in the real world produces billions of separately identifiable goods and services, covering at least several thousand product categories, I shall take the largest loss from the two three-sector models—that is, 3.2 percent—as the maximum estimate of the monopoly welfare loss.[7] This is considerably above the losses calculated by Harberger and Schwartzman, but considerably less than Bergson's maximum estimate and would fall further as more sectors are taken into account. Unfortunately, the calculating costs of considering anything like a realistic number of sectors are prohibitive.

APPENDIX

I shall now prove that *CNCV* is an increasing function of σ, for given monopoly price markups, and derive the formula for computing *CNCV* in Table 3. Using formula (3), define Y as follows:

$$Y^{\sigma-1} = \left[\frac{1 + CNCV}{\sum_k \left(\frac{\lambda_k}{\gamma_k}\right)}\right]^{\sigma-1} = \sum \gamma_k \lambda_k^{\sigma-1}$$

Assuming the $\{\lambda_k\}$ to be constant, and differentiating with respect to σ gives:

$$(\log Y)\, Y^{\sigma-1} = \sum_k \gamma_k (\log \lambda_k) \lambda_k^{\sigma-1}$$

whence:

$$\log Y = \frac{\sum_k \gamma_k (\log \lambda_k) \lambda_k^{\sigma-1}}{\sum_k \gamma_k \lambda_k^{\sigma-1}}$$

It will be sufficient to show that *log Y* increases with σ. We have:

$$\frac{d(\log Y)}{d\sigma} = \frac{\left(\sum_k \gamma_k \lambda_k^{\sigma-1}\right)\left(\sum_k \gamma_k (\log \lambda_k)^2 \lambda_k^{\sigma-1}\right)}{\left(\sum \gamma_k \lambda_k^{\sigma-1}\right)^2} - \frac{\left(\sum_k \gamma_k (\log \lambda_k) \lambda_k^{\sigma-1}\right)^2}{\left(\sum \gamma_k \lambda_k^{\sigma-1}\right)^2}.$$

But the numerator may be rewritten as:

$$\sum_{i>j} \gamma_i \gamma_j \lambda_i^{\sigma-1} \lambda_j^{\sigma-1} \left[(\log \lambda_i)^2 + (\log \lambda_j)^2 - 2 \log \lambda_i \log \lambda_j\right]$$

$$= \sum_{i>j} \gamma_i \gamma_j \lambda_i^{\sigma-1} \lambda_j^{\sigma-1} \left[(\log \lambda_i - \log \lambda_j)^2\right]$$

Given that at least one $\lambda_k > 1$, at least one $\lambda_k = 1$, and all $\lambda_k \geq 1$, the last sum above must be positive.

Turning to Table 3, we have, from formula (5):

$$\eta_1 = \frac{P_1}{MC_1} = \frac{\epsilon_1}{\epsilon_1 - 1} = \frac{m\eta_1}{m\eta_1 - 1}, \quad \text{since } \epsilon_1 = m\eta_1$$

$$\eta_2 = \frac{P_2}{MC_2} = \frac{\epsilon_2}{\epsilon_2 - 1} = \frac{p\eta_2}{p\eta_2 - 1}, \quad \text{since } \epsilon_2 = p\eta_2$$

By assumption, $\lambda_3 = 1$; $\gamma_1 = .33$; $\gamma_2 = .17$; $\gamma_3 = .5$. From (3) we have therefore:

$$1 + CNCV = \left[.5 + .33\left(\frac{m\eta_1 - 1}{m\eta_1}\right) + .17\left(\frac{p\eta_2 - 1}{p\eta_2}\right)\right].$$

---

[7] If there are additional intermediate goods producing sectors in our model economies whose own monopoly-price distortions alter costs of final-use goods producers and thereby raise total monopoly welfare losses by the same maximal percentage observed by Klass for the *U.S.* economy in 1958, the total maximal welfare loss rises to 4.5 percent (see fn. 2).

$$\left[\frac{1}{2}+.33\left(\frac{m\eta_1}{m\eta_1-1}\right)^{\sigma-1}+.17\left(\frac{p\eta_2}{p\eta_2-1}\right)^{\sigma-1}\right]^{1/\sigma-1}$$

$$=\left[1-\frac{.33}{m\eta_1}-\frac{.17}{p\eta_2}\right]\left[\frac{1}{2}+.33\left(\frac{m\eta_1}{m\eta_1-1}\right)^{\sigma-1}+.17\left(\frac{p\eta_2}{p\eta_2-1}\right)^{\sigma-1}\right]^{1/\sigma-1}$$

From formula (4):

$$\eta_1 = \sigma - .33(\sigma - 1) = .67\sigma + .33$$

$$\eta_2 = \sigma - .17(\sigma - 1) = .83\sigma + .17$$

Substituting these results into the above yields:

$$1+CNCV = \left[1-\frac{1}{(2.0303\sigma+1)m}-\frac{1}{(4.8824\sigma+1)p}\right]\cdot\left[\frac{1}{2}+.33\left(\frac{m(2.0303\sigma+1)}{m(2.0303\sigma+1)-3.0303}\right)^{\sigma-1}+.17\left(\frac{p(.83\sigma+.17)}{p(.83\sigma+.17)-1}\right)^{\sigma-1}\right]^{1/\sigma-1}$$

The formulas for the other tables are derived in identical fashion.

REFERENCES

A. A. Alchian, "The Basis of Some Recent Advances in the Theory of Management of the Firm," *Brit. J. Ind. Econ.*, Nov. 1965, *14*, 30–41.

W. J. Baumol, *Economic Theory and Operations Analysis*, 3d ed., Englewood Cliffs 1972, 325–27.

A. Bergson, "On Monopoly Welfare Losses," *Amer. Econ. Rev.*, Dec. 1973, *63*, 853–70.

W. S. Comanor and H. Leibenstein, "Allocative Efficiency, X-Efficiency, and the Measurement of Welfare Losses," *Economica*, Aug. 1969, *36*, 304–09.

A. C. Harberger, "Monopoly and Resource Allocation," *Amer. Econ. Rev. Proc.*, May 1954, *44*, 77–87.

M. Klass, "Inter-Industry Relations and the Impact of Monopoly," unpublished doctoral dissertation, Univ. Wisconsin, Madison 1970.

H. Leibenstein, "Allocative Efficiency, X-Efficiency, and the Measurement of Welfare Losses," *Amer. Econ. Rev.*, June 1966, *56*, 392–415.

J. Palmer, "The Profit-Performance Effects of the Separation of Ownership from Control in Large U.S. Industrial Corporations," *Bell J. Econ.*, Spring 1973, *4*, 293–303.

F. M. Scherer, *Industrial Market Structure and Economic Performance*, Chicago 1970, 400–09.

D. Schwartzman, "The Burden of Monopoly," *J. Polit. Econ.*, Dec. 1960, *68*, 627–30.

D. A. Worcester, Jr., "New Estimates of the Welfare Loss to Monopoly: U.S. 1956–60," *Southern Econ. J.*, *40*, Oct. 1973, 234–46.

On Monopoly Welfare Losses: Comment

By DEAN A. WORCESTER, JR.*

The long-standing belief that monopoly seriously misallocates resources, hurting consumer interest, received tremendously strong professional impetus in the writing of E. H. Chamberlin and Joan Robinson in the early 1930's. But statistical estimates by Arnold Harberger, David Schwartzman, and the author (1973) find losses trivial. Statistical studies by David Kamerschen and Frederick Bell show rather large losses, as much as 6 percent of national income, but seem to have defects (see the author, 1969).

Estimates showing small welfare losses have not satisfied many active scholarly experts studying industrial organization such as F. M. Scherer and William Shepherd, even less the regulatory authorities, Congressmen, radical economists, and the influential press. Recently, Abram Bergson reviewed Harberger's methodology, advanced an alternative procedure which avoids some flaws he finds in Harberger's approach, and published tables which display purely hypothetical ranges of welfare loss resulting from various assumed conditions. Bergson strongly implies that actual welfare loss is not "inconsequential" (pp. 864–68). But his more refined approach shares a basic defect with the above-mentioned researchers. All neglect rational behavior and fail to explore the relationship between firm demand and industry demand and supply. Thus Bergson's hypothetical tables display welfare losses for some combinations that cannot exist, and others that can exist only if monopolists choose values that grossly reduce their net income, even to large negative values.

This paper starts from well-known relationships between a firm's demand elasticity and profit-maximizing price, reviews the standard relationship between a firm's demand curve and the elasticity of its industry demand and supply curves, and proceeds to evaluate Bergson's and Harberger's efforts on the basis of what is known of statistical elasticities and margins of monopoly price above competitive price. Values that appear in Bergson's and Harberger's work are used throughout to facilitate exposition.

I. Elasticity, Monopoly Markups, and Welfare Loss

Estimates of elasticity of demand, or the elasticity of substitution, are critical to estimates of welfare loss because only the difference between competitive and monopoly price is subject to statistical determination. Output distortions are inferred from estimates of monopoly markups taken in conjunction with demand elasticity. The lower the elasticity, the smaller the output distortion and the smaller the welfare loss for a given monopoly price markup. Even small markups produce large welfare losses if elasticity is high enough.

Most statistically computed elasticities are small, often below unity, as displayed in Table 1. The zero and negative marginal revenues implied by unit and lower elasticities if relevant to individual firms are inconsistent with rational behavior. Hence, Harberger has been critized for using elasticities as low as -1.0. One of Bergson's innovations is to calculate welfare losses for elasticities of substitution as high as 64, which, for the situations he considers, involves elasticities of demand as high as -32.5 (p. 866).

Monopolistic competition theory implies that welfare losses are more properly based on demand curves facing individual firms, the elasticities of which may be very much higher than the computed elasticities in

---

* Professor of economics, University of Washington. Thanks are due especially to Yoram Barzel, and to Lowell Bassett, Judith Cox, John McGee, Tim Ozenne, Henry Pollakowski, Richard Parks, Potluri Rao, and George Stigler for helpful discussion and comments on key points. Josef Schwalbe searched the literature for statistical elasticities of demand following leads supplied by Parks and McGee. I, of course, remain fully responsible for errors of omission and commission.

TABLE 1—ESTIMATED ELASTICITIES OF DEMAND FOR INDUSTRIES

	Elasticities > 2.0			**Total = 8**	
Green Peas, fresh	2.8	(S)	Interest on Personal Debt, long run	2.7	(H)
Tomatoes, fresh	4.6	(S)	Intercity Railway, long run	2.7	(H)
Tomatoes, canned	2.5	(S)	Foreign Travel by *U.S.* Residents, short run	4.1	(H)
Medical Care and Hospitalization Insurance, long run	3.6	(H)	Airline Travel, long run	2.4	(H)
	Elasticity 1.0 to 2.0			**Total = 25**	
Apples, retail	1.3	(W)	Aluminum, at factory	0.4–1.2	(W)
Peaches, retail	1.5	(W)	Automobiles, retail	0.8–1.5	(W)
All fruit	1.1	(W)	Auto Insurance Premiums, Less Claims Paid, long run	1.6	(H)
Green Peas, canned	1.6	(S)	Street and Electric Railway and Local Bus	1.0	(H)
Watermelons	1.5	(S)	Nondurable Toys, long run	1.7	(H)
Beef, retail	0.8–1.1	(W)	Wheel goods, Durable Toys, Sports Equipment, Boats, and Pleasure Aircraft, long run	1.3	(H)
Pork, retail	0.7–1.0	(W)	Radio and Television Receivers, Records, short run	1.2	(H)
Lamb, retail	1.8–2.0	(W)	Radio and Television Receivers, Records and Musical Instruments, long run	1.3	(H)
Chicken, retail	1.2	(W)	Radio and Television Repair	1.0	(H)
Butter, retail postwar	1.3–1.4	(W)			
Shoe Cleaning and repairs, short run	1.1	(H)			
Shoe Cleaning and repairs, long run	1.5	(H)			
Laundering in Establishments	1.0	(H)			
China, Glassware, Tableware, Utensils, short run	1.2	(H)			
China, Glassware, Tableware, Utensils, long run	1.3	(H)			
Electrical Power for Residential Use	0.2–1.0	(W)			
(Selected Industries)	**Elasticities < 1.0**			**Total = 63**	
Cotton, at farm			Kitchen and Other Household Appliances		
Wheat, at farm			Steel, at factory		
Corn, at farm			Physicians		
Citrus Fruit, retail			Funeral and Burial Expenses		
White Potatoes			Legal Services		
Sweet Potatoes			Interest on Personal Debt, short run		
Eggs			New Cars, and Net Purchase of Used Auto Repair		
Milk			Gasoline and Oil, short run		
Margerine			Gasoline and Oil, long run		
All Feed					
Shoes and other Footwear					

Sources: (H) H. S. Houthakker and L. D. Taylor: (S) D. B. Suits. (W) L. W. Weiss.

Table 1, all of which are for industries.

There are two basic models that relate a firm's demand curve to that of its industry or group: a collusive and an independent pricing model. The collusive model supposes that each firm immediately matches the price of any other firm in the group, perhaps that set by a price leader. This, apart from nonprice competition, makes each firm's demand elasticity identical to the industry elasticity. Where this model applies, industry elasticities such as those presented in Table 1 are identical to firm elasticities and therefore are appropriate to the calculation of welfare loss, justifying Harberger's choice of unitary elasticity as a first approximation.

The independent pricing model is often expressed by a formula of the type:

$$e_f = \frac{1}{S} E_D - \left(1 - \frac{1}{S}\right) E_S$$

where e_f is the firm's demand elasticity, S its share of the market, E_D the elasticity of demand for its industry, and E_S the elasticity of supply of the rival firms. The same relationships can be shown diagramatically as in Figure 1, where we assume three firms with the same marginal cost curves, an industry demand elasticity of -1.0 and supply elasticity of 2.0 at the competitive price and output.

Under these circumstances the firm's demand elasticity is -7.0, and the welfare loss appears to be that represented by the large shaded triangle related to the firm's demand and marginal cost curves, and not the much smaller blacked out one related to the industry demand and supply curves.

Nevertheless, the -1.0 industry demand elasticity and the smaller triangle is relevant to the measurement of welfare loss. The reason is that the change of output Q_mQ_a by the monopolistic firm is partially offset by an opposite change, the amount AB, by its rivals, making the net change in the quantity of the good only Q_MQ_C. This amount can be found by applying ***industry*** demand elasticity to the margin between the competitive and the monopoly price, P_mP_c.

Figure 1 is identical to the diagram associated with the "dominant firm" which I have shown elsewhere to be unstable (1957). Depending upon cost conditions, the dominant firm either sacrifices quantity to maintain price by yielding market share to rivals until the firm loses influence, or, alternatively, its cost advantage is so great as to lead toward single firm monopoly. In the latter (rarely (observed) case, industry elasticity obviously becomes relevant. In the former, welfare loss is progressively lessened. Each firm's elasticity rises, approaching as S approaches zero the infinitely high elasticity of firm demand under pure competition. It follows that welfare losses tend to diminish over time in the absence of innovation that creates new short-run monopolies.

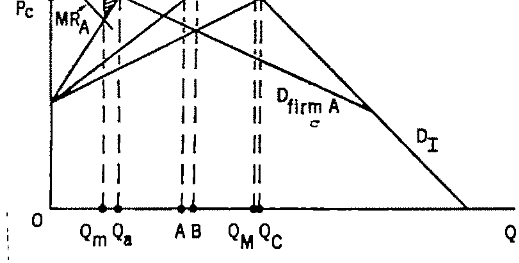


FIGURE 1. FIRM AND INDUSTRY ELASTICITY

The welfare loss calculated with industry elasticities appears to be the upper limit also in the short run. The still controversial sales maximization hypothesis would lead one to expect monopoly markups lower than the profit-maximizing ones where managers have the options that monopoly confers. Other motives such as the desire to run a larger enterprise, achieve financial advantages from size, hold market share in an expanding market, and perhaps price so as to discourage entry may also be served by lower margins. I also found maximizing strategy against uncertainty to involve lower margins (see the author 1967, especially p. 102). All of these require larger-than-monopoly outputs and smaller welfare losses.

I must conclude that not only are industry, not firm, elasticities relevant to the measurement of welfare losses, but also that industry elasticities overstate the size of loss.[1]

---

[1] Although industry elasticity is relevant for the calculation of welfare loss, elasticity by brand is of interest because it provides a basis for choice among alternative theories relating firm to industry demand. Keith Cowling recently calculated demand elasticities by brand for British cars, tractors, margarine, coffee, and toothpaste in order to isolate the effects of advertising. Quality differences were also estimated for cars. Quantity changes are given exclusively as market shares of industry sales. The elasticity of industry demand is not clearly stated (p. 101, n). The following tabulation presents Cowling's short-run demand elasticities (col. 1), and, where available, the industry elasticities for comparable *U.S.* industries. The brand elasticities are considerably larger and therefore inconsistent with the collusive price matching theory. Cowling declares his short-run estimates to be "much smaller" than the long-run ones (p. 101). For example, the long-run elasticity for British cars is given as -7.06. But there are statistical difficulties involving lags.

	(1)	(2)
Cars (*OLS*)	-1.950	-0.8 to 1.5 retail (H)
(*TSLS*)	-1.915	-0.6 short run (S)
Tractors mean	-3.290	—
1965	-4.630	—
Margarine (*OLS*) annual	-4.032	-0.3 (W)
Coffee (*TSLS*)	-0.202	—
Toothpaste (*OLS*)	-1.977	—

Sources: Cowling, p. 102; (H) Houthakker and Taylor; (S) Suits; (W) Weiss. The brand elasticities are, nevertheless, smaller than reasonably expected from the independent pricing theory.

II. What Values Can Properly Be Given to the Parameters?

The size of welfare loss depends importantly upon the size of the monopolistic sector, the extent to which monopolistic prices exceed competitive levels, the distribution of monopoly markups, and the size of the relevant elasticities. Definitive studies establishing these values do not exist, but it is possible to rule out wide ranges of values

The effect of the size of the monopolistic sector, however, is ambiguous on its face. As Joan Robinson pointed out long ago, a "world of monopolies" would produce exactly the same amount of each commodity as would a "world of competition," if each confronted the same (greater than unitary) demand elasticity. Only income distribution between marginal and residual claimants would be affected (ch. 27). I disputed even that (1967, pp. 199–208). But if monopoly power is said to exist only when a firm manages to hold its price consistently 5 percent or more above that needed to earn an average rate of return on investment, less than 6 percent of the firms covered by the *Fortune 500* survey qualified during the fourteen-year period ending with 1969 (see the author, 1973, p. 243).

The extent to which firms succeed in obtaining prices that produce high profits usually[2] affects the size of welfare loss, but to a lesser extent than might be expected. Data showing total net income as a percent of sales rarely suggest markups as high as 10 percent. A simple average of Harberger's industries has average monopoly price 1.072 of the average in the competitive sector. Suspicious investigators therefore look for profits hidden one way or another in costs. Rents, royalties, inflated executive emoluments, advertising, and inefficiency coincident with the choice of a quiet life are suspected hiding places. The alleged monopolist's margin over competitive price is then inflated on the implicit assumption that no function valuable to consumers is performed by these expenditures. But this view is weakening as writers such as Lee Benham and George Stigler (1961), to name only two, find advertising to be valuable to consumers as a source of information and a promoter of competitively lower prices; and as environmental economists rediscover the importance of rents to the allocation of resources, indeed to the very preservation of assets that yield rents.

Still, high margins can be generated. Kamerschen made a detailed effort to estimate the size of monopoly profits that may be hidden in categories not reported as profit for each of sixty industries. His data cover corporate and noncorporate sectors adding, in addition to reported profits in excess of average profits, all of advertising expense, royalties, and such intangibles as the return to patents, trademarks, and good will. He believes that he has also included rents. No allowance is made for profits hidden in executive compensation or wages, but neither was any effort made to exclude short-term profits that would appear in a fully competitive world during periods of adjustment to changed demand and supply conditions. Although Kamerschen recognizes as extreme the allocation of all of advertising to monopoly profit, he believes that on balance he underestimates the size of monopoly markups.

The size of his markups is not published, but its average size can be estimated by comparing the size of the welfare losses when he uses an elasticity of -1.0 to those when he uses a "Lerner" elasticity calculated from the standard equation for elasticity, $P/(P-MC)$, where in this case P is the monopoly price and MC is the competitive price calculated as P minus his estimate of monopoly profits per unit of output.

Kamerschen's choice from the thirty-two pairs of welfare losses displayed in his Table 1 (p. 230–31) shows a welfare loss of 6.11 percent of national income for the computed elasticity, and 1.87 for unit elasticity. Since the welfare loss at higher elasticities is the loss under unit elasticity times the higher elasticity, this yields an implied elasticity of -3.26, which, if $P_c = 1.0$, produces a monopoly markup of 44 percent of competitive price or 30.5 percent of actual price. It is difficult to reconcile this markup with national account data that show compensation of employees in the corporate sector claiming 66 percent of gross product originating in nonfinancial corporations.[3]

---

[2] An exception relates to decreasing cost industries that may earn high profits by price discrimination that markedly improves resource allocation. This is an aspect of the typical assumption that all industries operate at constant cost. Where rising cost is the fact, welfare loss is also overstated.

We leave the matter of size of monopoly markups unresolved, but with considerable skepticism attached to values which require net income above competitive levels to exceed 15 percent of sales for more than a few years.

More interesting is the fact that this high markup, taken in conjunction with the profit-maximizing assumption that makes it determine the elasticity of demand used to compute welfare loss, produces a loss of "only" 6.11 percent of national income. If the industry elasticities displayed in Table 1 were used in connection with Kamerschen's markups, as is appropriate, welfare loss would be considerably below the 1.87 percent of national income he gives for unit elasticity. This points up the importance of the third variable that affects the size of welfare loss, elasticity.

It is evident that the −3.26 elasticity computed from Kamerschen's high monopoly markup is much higher than those found in Table 1. Sixty-three of the ninety-six industries have elasticities less than unity, and the lower end of ranges given for five more fall below −1.0. These imply negative marginal revenue curves, and monopoly markups larger than the total price if each firm enjoys a market share demand curve. An additional twenty-five would have markups of 100 percent to 50 percent of price if their industries were monopolized: Some economists have argued that such data imply that the firms are not profit maximizers, and for various reasons produce more than the profit maximizing amount. An argument more in accord with economic analysis is that they are not monopolists.

Examination of the second theory relating firm demand elasticity to industry elasticities of demand and supply suggests why we should expect to find what is usually found, relatively small markups by profit-maximizing oligopolists despite very low industry elasticities. An oligopolistic firm's demand curve is much more elastic than its industry's if they do not collude. For example, apply the equation for independent pricing to the second largest *U.S.* oil company which has about 5 percent of industry sales. If the industry demand elasticity is a reported −0.46 and the elasticity of other firms' supply is assumed to be a very inelastic 0.25, the firm's demand elasticity comes to −13.95. Pricing to maximize profit with this demand elasticity suggests a margin about 7.7 percent above the competitive level. An attempt to establish a larger markup would insanely reduce profits. Smaller industry shares and larger supply responses by other firms drastically lower the margin. If the elasticity of supply is 1.0 instead of 0.25, the firm's demand elasticity is −28.2 and the profit maximizing markup over competitive price only 3.67 percent. The same figures for the largest *U.S.* oil company are approximately −13.6 and 7.9 percent. But the largest firm has no assurance that the others will not increase capacity and output more rapidly should it maintain the higher price for a period of time.

One may doubt the significance of such high coefficients. After all, firms are subject to the same general influences and tend to move their prices in the same direction at about the same times. Thus high elasticities are rarely if ever observed. But such argument misses the point. Competitive firms move their prices more or less together in response to commonly felt market influences. Statistical demand curves as a result are relevant to generic products, not the outputs of individual firms. High elasticities are relevant to independent behavior. Remorseless discipline constrains firms attempting to raise a price for reasons other than the general market influences felt by others.

---

[3] One puzzle is the size of the welfare loss, 1.57 percent of national income, found by Kamerschen when unadjusted profit data and unit elasticity are used. The assumptions would seem to accord with Harberger's assumptions and procedures which produced a welfare loss of only 0.08 percent. The most likely explanation of the difference involves Kamerschen's inclusion of the noncorporate sector which may substantially widen the range of return around the average. Still, the estimate seems high.

In principle, collusion offers benefits to each firm in an industry. But a firm with only 5 or 10 percent of the market can hope to gain by surreptitious price cuts not easily detected or policable by rivals.[4] Private oligopoly has no enforcement power comparable to legal prices backed by civil and criminal governmental sanctions. Even those are imperfectly enforceable.

According to the statistical record the vast majority of profitable large firms operate on relatively narrow margins, securing prices that return less than 5 percent on sales above that necessary to earn average profits. This implies firm demand elasticities larger than twenty despite inelastic or very modestly elastic industry demand curves, such as those calculated from statistical data. Armed with these tools and data we can analyze Harberger's and Bergson's studies of welfare loss.

III. Comparison of Harberger's and Bergson's Findings

Harberger applies a 1938 model developed by Harold Hotelling to Eptstein's data for seventy-three *U.S.* industries, 1924–28. Deviations from average profits divide the group approximately evenly into "monopolistic" and "competitive" sectors. The deviations of price from the competitive level are calculated from profit data, competitive prices being "too low" because on the relevant full employment assumption, overexpansion of the competitive sector follows reduced opportunities in the monopolistic sector where, of course, prices are "too high." Harberger assumes a universal unitary elasticity of demand to make possible a statistical estimate of welfare loss which turns out to be 0.08 percent of national income.

Bergson's approach is more elegant. He uses a constant elasticity of substitution (*CES*) utility function and a linear production function to calculate the compensating variation—the additional income consumers must receive to offset the loss of utility occasioned by monopoly pricing. The share of the total market monopolized, the size of monopoly price markup over competitive price, and different elasticities of substitution (used in place of demand elasticity), are built into his model. Potential welfare losses are found by assuming specific values for the parameters.

Bergson does not make estimates drawn from statistics. Instead, two tables are offered displaying welfare losses calculated for a number of hypothetical alternatives. In every case, half of the economy is competitive. Monopoly price is assumed to be alternatively 10, 20, and 30 percent above the competitive price. The elasticity of substitution is postulated, alternatively, as 0.5, 1.0, 2.0, 4.0, 8.0, 16.0, 32.0, and 64.0 (pp. 864–65). The relationship between elasticity of substitution and elasticity of demand given (p. 866) makes the equivalent demand elasticities the negative of 0.75, 1.0, 1.5, 2.5, 4.5, 8.5, 16.5, and 32.5 in this case. Welfare loss is highest where the monopoly markup is 30 percent and the demand elasticity is -32.5.

Since the object of Bergson's effort is clearly to provide theoretical support for the ancient belief that monopoly may seriously injure consumers, it is important to see if his model produces significantly higher estimates of welfare loss than others, such as Harberger's, when data gathered by others are plugged in. This is easy to do. Harberger's method also divides industry about 50 percent to monopoly, his deviations of returns on sales show average monopoly price to be 7.2 percent above the competitive price, and he assumes a unitary demand elasticity.

The natural comparison is to Bergson's result for monopoly prices 1.1 times the competitive level, calculated for a unitary elasticity of substitution equal to demand elasticity in this instance. The loss shown in Bergson's Table 1 (p. 864), 0.11 percent of national income, is about as inconsequential as the 0.08 percent found by Harberger for monopoly prices 1.072 times the competitive level.[5]

---

[4] For an acute analysis of this problem, see Stigler (1964).

[5] Nor is the effect of higher elasticities under Harberger-type analysis significantly different from Bergson's as long as monopoly prices are only 1.1 of competitive prices. The Hotelling analysis used by

Bergson finds a larger loss when the monopoly sector is divided into two parts, one with substantially higher prices than the other. Whether or not it becomes "consequential" depends partly on how big the loss must be to deserve that label, but it should be noted that the highest loss Bergson gives for unitary elasticity, the elasticity used by Harberger, is 1.29 percent of national income. That figure is based on the assumption that 17 percent of consumer expenditures are for products with prices 60 percent above the competitive level and 33 percent for products priced 15 percent above the competitive level (p. 865, Table 2, col. 6).

As can be observed from Bergson's tables (pp. 864–65), there are no truly large welfare losses in the absence of large elasticities. Only large markups in conjunction with extremely high elasticities of substitution, as high as 64, equivalent to -32.5 demand elasticity, produce values for monopoly loss that are certainly "consequential"—as high as 39.03 percent of national income. His practical conclusions therefore, rest most heavily upon the size of elasticities in actual markets.

Table 1 presents all the statistical elasticities of industry demands that a determined search of the available materials turned up. As may be seen, elasticity is above 2 in only eight instances, none in industries that absorb a large proportion of consumer income. Virtually every important industry has a demand elasticity less than -1.5, far less than the level required for consequential welfare losses in Bergson's tables even for monopoly markups as extreme as 30 percent for half of *U.S.* output, if welfare losses must exceed 2 percent of national income to be deemed consequential.

A much more serious objection must be raised against Bergson's tables. Most of the combinations of monopoly markup and elasticity are entirely ***inconsistent with profit-maximizing behavior***, and would indeed require firms to raise prices far ***above*** the profit-maximizing levels. When these combinations of elasticity and monopoly prices are omitted from Bergson's tables, maximum welfare losses ranged between approximately 2 and 4 percent.[6]

Even worse, the implied competitive equilibria calculated from combinations of markup and elasticities that produce high losses require large ***negative*** outputs in the competitive sector. This can be demonstrated with a simple constant cost numerical example. Let full employment income equal \$2.6, half spent at a competitive price of \$1, and half at a monopoly price of \$1.3. Total factor earnings are \$2.3; monopoly profits are \$0.3. Let price in the monopoly sector be cut to \$1, a reduction of 23 percent. We are at liberty to pick from Bergson's tables the equivalent of an elasticity of demand as high as -32.5. Suppose it is -30. The result is a 690 percent

---

Harberger allows one to calculate the effect of larger elasticities of demand by multiplying the welfare loss when elasticity is -1.0 by alternative demand elasticities (Harberger, note 2, pp. 81–82). Bergson's highest estimates are for an elasticity of substitution of 64. The demand elasticity equivalent is -32.5. Multiplying the 0.11 percent welfare loss when the demand elasticity is -1.0 by 32.5, we have 3.575 percent of national income. Bergson's method yields 3.86. Bergson's values diverge substantially from Harberger's when monopoly prices are 1.2 and 1.3 times the competitive level, but not in the direction favorable to Bergson's conclusion. He finds losses of 8.80 and 13.74 percent of national income when the elasticity of substitution is 64 (as compared to .41 and .86 when it is 1.0). The Hotelling-Harberger approach yields much larger values, 13.32 and 27.9 percent, respectively.

[6] These maxima require the assumption, seemingly popular in some quarters, that monopolists are malevolent, out to maximize harm to consumers regardless of cost to themselves. Monopolists can accomplish this either by raising their prices until the consumers can afford none of the monopolist's products, or by lowering price until the competitive sector is eliminated. The monopolist can price himself out of the market with price increases of either 20, 10, 5, or 2.5 percent as the demand elasticity is alternatively 5, 10, 20, or 40 ($\Delta Q/Q = -100$; $\Delta P/P = 100/E$). A reduction of sales of the initially competitive sector to zero involves lowering price enough to double total dollar sales, leaving nothing for the competitive sector. Reductions of .1, .2, and .3 used in Bergson's examples are percentage reduction of 9, 16⅔, and 23 percent. Since in each case the limits of production possibility are reached when monopoly output is doubled if true marginal costs equals competitive price 1.0, elasticity can be found by dividing the percentage change of price into the required 100 percent change of output. The elasticities are -4.35 when the monopolist's price reduction is from 1.3, -6.0 when it is from 1.2, and -11 when it is from 1.1. Note that the larger monopoly margins are associated with the smaller elasticities.

increase in output in the monopolized sector, to an output of 6.9. Under constant costs this requires a transfer of resources valued at \$5.9 to the formerly monopolized sector. This is \$4.6 more than the total employed in the original competitive sector. Thus, elimination of monopoly implies an impossibility, a large negative output in the original competitive sector.

Actually, the profit-maximizing markup when elasticity of demand is -30.0 is only 3.45 percent.

IV. Conclusions

Welfare loss literature suffers from the failure to apply standard propositions from the theory of the firm to the relationships between elasticity, marginal revenue, and profit-maximizing prices, and neglect of the relationship between the elasticity of a firm's demand curve and the elasticities of supply and demand of the industry of which it is a part. Bringing them in provides a rational explanation for the paradox of modest markups by profit-maximizing oligopolists in industries with inelastic demands. Standard theory also leads to the conclusion that it is correct to use industry elasticity of demand (not the firm's) to calculate welfare losses. When translated into the appropriate elasticities of substitution and applied to profit-maximizing markups on sales in the corporate sector, Bergson's procedure, more elegant than Harberger's, yields essentially his result. Nothing in Bergson's article provides any basis to believe in the relevance of markups or elasticities larger than those found by previous researchers. Examination of Bergson's case, therefore, does not controvert, but rather reinforces the hypothesis that monopoly losses are "inconsiderable."

It should be added that estimates of moopoly loss in labor markets by Bell and Albert Rees, not referred to by Bergson, find welfare losses somewhat larger than in consumer markets but still "small" especially, in the case of Bell, when welfare loss is confined to net changes from reallocation.

No doubt some true misallocation occurs in factor markets, thus shrinking the production possibility curve. The extent to which the value of the misallocated resources surface in profits and can be separated from functionally useful profits has not as yet been determined.

The time is surely close at hand for economists to accept the research findings that welfare losses due to enterprise monopoly are small in the American economy. On the evidence, industry approximates the competitive model, ancient and conventional wisdom, playwrights, publicists, John Kenneth Gailbraith, and many lucky business leaders to the contrary notwithstanding. The appearance of discretionary power is not the same as power. Business leaders, like all competitors, may typically occupy the position of the Red Queen, who had to run as hard and intelligently as she could to stay more or less in place, although apparently a royal despot and red at that.

REFERENCES

L. Benham, "The Effect of Advertising on the Price of Eyeglasses," *J. Law Econ.*, Oct. 1972, *15*, 337–52.

F. W. Bell, "The Effect of Monopoly Profits and Wages on Prices and Consumer Surplus," *Western Econ. J.*, June 1968, *6*, 233–41.

A. Bergson, "On Monopoly Welfare Losses," *Amer. Econ. Rev.*, Dec. 1973, *63*, 863–70.

E. H. Chamberlin, *The Theory of Monopolistic Competition*, 1st ed., Cambridge, Mass. 1933.

K. Cowling, "Optimality in Firms' Advertising Policies and Empirical Analysis," in his *Market Structure and Corporate Behavior*, London 1972.

A. C. Harberger, "Monopoly and Resource Allocation," *Amer. Econ. Rev. Proc.*, May 1954, *44*, 77–87.

H. Hotelling, "The General Welfare in Relation to Problems of Taxation and Utility Rates," *Econometrica*, July 1938, *6*, 242–49.

H. S. Houthakker and L. D. Taylor, *Consumer Demand in the United States: 1929–1970*, Cambridge 1966.

D. R. Kamerschen, "An Estimation of the 'Welfare Losses' from Monopoly in the

American Economy," *Western Econ. J.*, Summer 1966, *4*, 221–36.

A. Rees, "The Effects of Unions on Resource Allocation," *J. Law Econ.*, Oct. 1963, *6*, 69–78.

J. Robinson, *The Economics of Imperfect Competition*, London 1934.

F. M. Scherer, *Industrial Market Structure and Market Performance*, Chicago 1970.

D. Schwartzman, "The Burden of Monopoly," *J. Polit. Econ.*, Dec. 1960, *68*, 627–30.

———, "The Effect of Monopoly: A Correction," *J. Polit. Econ.*, Oct. 1961, *69*, 494.

W. G. Shepherd, *Market Power and Economic Welfare*, New York 1970.

G. J. Stigler, *Theory of Price*, 3d ed., New York 1966.

———, "A Theory of Oligopoly," *J. Polit. Econ.*, Feb. 1964, *72;* reprinted in his *The Organization of Industry*, Homewood 1968.

———, "The Economics of Information," *J. Polit. Econ.*, June 1961, *69*, 213–25.

D. B. Suits, *Principles of Economics*, New York 1970.

L. W. Weiss, *Case Studies in American Industry*, New York 1967.

D. A. Worcester, Jr., "Innovations in the Calculation of Welfare Loss to Monopoly," *Western Econ. J.*, Sept. 1969, *7*, 234–43.

———, "New Estimates of the Welfare Loss to Monopoly: U.S. 1956–69," *Southern Econ. J.*, Oct. 1973, *40*, 234–46.

———, *Monopoly, Big Business and Welfare in Postwar United States*, Seattle 1967.

———, "Why Dominant Firms Decline," *J. Polit. Econ.*, Aug. 1957, *65*, 338–46.

On Monopoly Welfare Losses: Reply

By Abram Bergson*

I

In his very interesting note, Richard Carson undertakes to adapt the illustrative calculations on monopoly welfare losses in my 1973 article, so as to allow for certain relations among the underlying parameters. The relations in question, he believes, are indicated by the classical theoretic condition for monopoly pricing.

$$p/MC = \epsilon/(\epsilon - 1) \tag{1}$$

My illustrative calculations were intended to be just that, and from my own discussion of (1), pp. 861, 867, the reader, I think, will not be very surprised if losses are reduced when account is taken of that condition. Whether they must be as limited as Carson argues, however, is another matter. In any event, I do have misgivings about his computations, and hence also about his resultant "maximum estimate."

In making our calculations, both Carson and I apply a formula for monopoly welfare losses (my (18), Carson's (3)) that I derived on the assumption that in the economy considered there is but a single consuming household, though one that acts as a perfect competitor. Also, the household's indifference map is given by a *CES* function. As it turns out, losses then depend on the elasticity of substitution (σ) in that function, and on the magnitudes, for all products, of two further variables: λ_i, the ratio for the ith good of price to marginal cost, and γ_i, the share of expenditures on that good in the household's income. Note also that η_i the own-price elasticity of demand for the ith good, varies with σ and γ_i in conformity with a further formula that I derived: my (22), Carson's (4). Hence, for any given σ and any given magnitudes of γ_i for different products, there is a corresponding set of magnitudes of η_i.

While we thus take the same formula as a point of departure, Carson and I diverge in applying it. As implied, the divergence turns on the use made of (1). Thus, for purposes of my illustrative calculations, I in effect treated σ, λ_i, and γ_i as independent parameters, and allowed each in turn to vary over what seemed an appropriate range without regard to any constraint that might be imposed by (1). By implication, despite (1), I also allowed η_i and λ_i to vary independently.

Carson too takes σ as a parameter, and also allows γ_i to vary much as I did. As for λ_i however, that is given throughout by (1), on the understanding that ϵ is the "... demand elasticity relevant to a particular price-setting decision maker within the industry" (Carson, p. 1010). Also, several alternative hypothetical relations are entertained between ϵ and η_i the demand elasticity that is immediately relevant to monopoly welfare losses in my (22) (Carson's (4)): in an economy with two sectors, one of which is monopolistic, ϵ is taken in turn to equal η_i; 2 η_i, and 3 η_i.[1] In an economy with three sectors, two being monopolistic, each of the foregoing three relationships obtains in turn in each monopolistic sector, though the same relationship need not obtain simultaneously in both of them. What is to be said for this procedure as an alternative to mine?

In trying to answer, it should be observed first that the model that both Carson and I use is not entirely satisfactory for purposes of formally integrating (1) into the calculations. At least, what one might wish to allow for in that case is any resultant cross-sectoral covariation of η_i and λ_i over the relevant range. Given a *CES* indifference map, however, σ is necessarily uniform for all pairs of commodities. While η_i may still vary between commodities, the variation according to my (22) (Carson's (4)) necessarily is limited and awkwardly can occur only in dependence on the variation in γ_i. The latter, to repeat, is the share of the ith good in the household's budget.

---

* Professor of economics, Harvard University.

[1] Here and elsewhere, I follow Carson notationally in understanding that in relations such as are in question, ϵ like η_i pertains to the ith product; that is, ϵ pertains to the ith product, so far as (1) applies to that product. On whether the formula does so apply, see below.

While making illustrative calculations for a *CES* indifference map, however, I also developed formulas for monopoly welfare losses—(1), (5), and (8)—that apply more generally. Although relatively complex, these fairly clearly indicate that the losses must depend on, among other things, η_i and λ_i. Also, other things equal, the losses as might be expected must be greater the larger the η_i for a given λ_i and the larger λ_i for a given η_i. It was in the context of this more general model that I ventured to speculate (as I suggested above that I did) as to the possible import of (1) for monopoly welfare losses. We may also, I think, properly construe Carson in the same light. While losses are calculated within the framework of the *CES* model, the concern evidently is somehow to represent the impact of cross-sectional relations between η_i and λ_i such as might be indicated by (1).

Though evident, it is advisable to underline that η_i, the own-price elasticity of demand for a product, is understood in the usual way in terms of a demand schedule such as:

$$(2)\quad x_i = d^i(p_1, p_2, \ldots, p_i, \ldots, p_n, I)$$

The own-price elasticity of demand for x_i, then, is the negative of $(p_i/x_i)(\delta x_i/\delta p_i)$, where in the partial derivative all p other than p_i are constant; and so too, of course, is I, the household's income. Also, a product is understood as an item sold by a firm for which the ratio of price to marginal cost differs from the corresponding ratios for other items sold by either the same firm or other firms. Hence, as I state in my original article:

> . . . what counts is variation [in the ratio of price to marginal cost] not only between industries in any conventional sense, but between products; even between different qualities or models of the same product (for example, different models of Buicks) so far as price-cost ratios vary here as well. [p. 860]

These interpretations of η_i and of the concept of a product obtain as these aspects relate to the calculation of monopoly welfare losses under the *CES* model. They also obtain, however, where reference is to losses more generally.

To come back to Carson, evidently one question at issue concerns the principles of monopoly price formation. That is also a matter that is hardly to be settled here, but in applying (1), Carson, to repeat, takes ϵ to be the "demand elasticity relevant to a particular price-setting decision maker within the industry." On that understanding, I don't think there is any need to quarrel especially with his view of monopoly price formation in order to conclude that in his calculations he allows for some, but by no means all, cases of monopoly price formation that are of interest. In order to see that that is so, it suffices to recall pertinent theoretic precepts for the three main textbook sorts of monopolistic markets:

(i) *Monopolistic Competition*; i.e., where an "industry" consists of many small firms each producing a good similar to, but yet differentiated from, those produced by other firms, and where each firm fixes its price quite without regard to any impact its actions might have on its competitors. Here (1) must indeed hold in equilibrium, or so at least the primers teach us. In practice, there are the usual caveats as to business motivation and knowledge of costs and demand, but Carson claims no more for (1) than that there be a tendency for it to hold and I think we must grant a presumption to that effect. It also follows that λ_i should be more or less in conformity with (1), and that ϵ should approximate η_i. In short, here Carson's premises must be granted, at least if the primers are at all trustworthy. Indeed, the first of his three hypothetical relations between ϵ and η_i, that where the two elasticities are equal, is essentially the one that obtains.

(ii) *Pure Monopoly*; i.e., a single firm produces all of a product for which there are no close substitutes, and without there being any threat to speak of from potential rivals. The indicated conclusions regarding (1), λ_i, ϵ, and η_i are, of course, as in (i), so here too Carson is essentially validated, and with ϵ for the product in question equal to the corresponding η_i.

(iii) *Oligopoly*; i.e., the industry consists of relatively few firms producing more or less similar though not necessarily identical products. Here unfortunately, the primers, and indeed even the advanced texts, are not always as instructive as one might wish. But η_i for the product of any one firm manifestly might frequently be very large. In fact, depending on how similar the products of different firms are, η_i might even approach infinity. On the other hand, oligopolistic firms seem characteristically to avoid any very consequential price competition. Hence, in fixing the price of its own product, the individual firm is more or less concerned with the response of demand not only to a change in that price but to possible concomitant changes in the prices of similar products of his competitors. To what extent in such circumstances λ_i might still be explained by a simple formula such as (1) is an interesting question, but so far as it is the relevant ϵ clearly must often be less than η_i and possibly far less. In any event, λ_i should often be greater, and sometimes much greater, than (1) would indicate with ϵ taken to equal η_i.

At this point, then, Carson is surely wide of the mark. Thus, for oligopoly, one wonders whether there can be any systematic relation to speak of between λ_i and η_i. So far as there is, it evidently can hardly be such as Carson obtains when in applying (1) he improperly allows ϵ to equal η_i. The incongruity between his λ_i and η_i necessarily only becomes the greater in his alternative cases where ϵ is taken to equal 2 η_i and $3\eta_i$.

Why then does Carson delineate the domain of his parameters in that way? Apparently, his rationale is found in calculations indicating that, contrary to a finding of mine, disaggregation reduces calculated losses. Hence, in order to obtain an upper limit for losses, we may safely think of η_i as referring to a composite for broad sectors, rather than as I have urged is necessary, for finely differentiated products. By the same token, ϵ may be taken as being, if anything greater, rather than (as I have argued) less, than η_i. Or so Carson appears to reason, but I don't think properly.

We must distinguish two approaches to aggregation. In both, reference is to composites—to be precise, let us say Hicksian composites—for groups of distinguishable but related products. Under one approach, however, the price of each composite itself is supposedly set by a single hypothetical monopolist. Hence, the price-cost ratio for the composite conforms to (1), with ϵ taken as the elasticity of demand for the composite, or some multiple of that elasticity deemed typical for the market in question. Under the other approach, the price-cost ratio for the composite is simply taken to be the average of the price-cost ratios for individual products comprising it. Whichever the approach, the question is posed as to how calculated monopoly welfare losses are affected if in the computation, rather than consider the price-cost ratios for individual products in each composite, we allow the price-cost ratio for the composite to stand for all of them together. The answer to that question, however, evidently could differ depending on which of the two sorts of price-cost ratios for composites is referred to.

Of the two approaches, the second seems essentially the one applied in empirical research such as that in the pioneer paper of Arnold Harberger. Thus, Harberger focuses there on price-cost ratios that are in effect sectoral averages and have no relation to the single ϵ, unity, which is assumed to prevail everywhere. On the other hand, David Kamerschen in one of several computations does assume that (1) applies at a sectoral level. Curiously, however, that formula is used to estimate elasticities rather than price-cost ratios. The latter here, too, are in effect sectoral averages.

The difference between Carson and me regarding aggregation must be seen in the foregoing light. Of the two approaches, Carson clearly is concerned with that where a single hypothetical monopolist fixes the price of each composite. That approach, at any rate, underlies his computations; his exposition might sometimes suggest that he is proceeding otherwise. Carson's approach is of interest, but the results have no bearing one way or other on my finding that with disaggregation calculated losses increase, for I focus on the alternative approach where the price-cost ratio for each composite is simply an average of the price-cost ratios of the corresponding individual products. In applying that approach, moreover, I properly allow for divergencies in such price-cost ratios within any composite. In contrast, as we may infer, of six cases of disaggregation that Carson considers as vindicating his thesis, (i.e., two cases for each of the hypothetical relations allowed between ϵ and η_i), three are such that λ_i is the same for two subsectors in question. In all six cases, the λ_i for both subsectors are less than the price-cost ratio for their composite. I come in a moment to the reasons why Carson's price-cost ratios turn out to be restricted in these ways. Suffice it to observe now that with the ratios so restricted, none of Carson's six cases comes within the ambit of the kind of disaggregation on which I focused.

II

But granting the difference in approach, does it not remain true that for Carson disaggregation reduces calculated losses? If so, is he not warranted in thinking that, for purposes of deriving an upper limit for losses, my (18) (his (3)) may properly be construed as relating to composites; and hence that his calculations, when construed similarly, are illustrative of losses under monopoly generally, including oligopoly, even though he takes $\epsilon \geq \eta_i$? In a word, does he not by using his approach arrive after all at an upper limit for losses, as he seeks to do?

That seems more than doubtful when we consider further the six cases which Carson takes as demonstrating that losses decrease with disaggregation. As indicated, in all these cases sector and subsector price-cost ratios are related in rather special ways. There are two reasons for that. One is simply that Carson follows me in using a *CES* indifference map. Given that, the resultant need to conform to my (22) (Carson's (4)), and also Carson's imposition of (1)—given all that, as readily seen, Carson's sector and subsector price-cost ratios had to be rather special. Possibly for consumer's tastes more generally, disaggregation of the Carson kind would reduce losses but if so, that fact has still to be demonstrated.

The other reason why Carson's sector and subsector price-cost ratios turn out to be of very restricted sorts is that in disaggregating Carson focuses exclusively on "comparable situations." That means that ϵ in (1) is the same multiple of η_i in the two subsectors considered, and that for the corresponding sector it is also that multiple of the sectoral demand elasticity. Those restrictions probably pose no special difficulty where reference is to monopolistic competition and pure monopoly. In such markets equation (1) tends to obtain with ϵ generally approaching η_i for different products. On the other hand, under oligopoly as we saw so far as (1) applies at all, ϵ tends to vary in relation to η_i within a product group, and often markedly. To the extent that ϵ does so vary, therefore, the reduction of losses with Carson's kind of disaggregation remains unproven under oligopoly even with a *CES* indifference map.

Actually, the demonstration of that result can not be considered as extending to oligopoly in any event since cases where $\epsilon < \eta_i$ for individual products are illegitimately excluded even for the purposes of that exercise. But at this point it must be acknowledged that Carson himself is needlessly restrictive. While oligopoly must often entail $\epsilon < \eta_i$, as is easy to see, even with that relation disaggregation such as is in question still would reduce losses; provided, that is, that reference is still to a *CES* indifference map and to "comparable situations."[3]

I have been tacitly assuming that Carson's

---

[3] Although Carson refers only to cases where $\epsilon \geq \eta_i$ for subsectors, we may be able to gain some insight into the import of the varying relation of ϵ to η_i for individual products under oligopoly by juxtaposing some of Carson's calculations for situations that are not "comparable." Of particular interest I think are these cases: in Carson's notation $m=2$ for a single monopoly sector, and $m=1$ and $m=3$ for two subsectors. The sectoral multiple is thus the mean of the two subsectoral multiples, which seems appropriate under Carson's approach to aggregation. As seen from Carson's tables, disaggregation here markedly increases losses.

composites refer to product groups of a more or less conventional sort. That appears to be his supposition too. But what if attention were focused instead on very large segments of the economy, or even on one segment embracing each and every monopolized product produced? Might not the losses inflicted by the corresponding hypothetical monopolists be an upper limit for actual losses, at least in these cases? In fact, is that not quite possible whatever the nature of the indifference map and whether situations are comparable or not? No doubt it is, but the upper limit could be high indeed. To be aware of that, one need only ponder for a moment the possible impact of a single monopolist controlling, say, a country's entire food supply.

Whatever the scope of the composites, note that with Carson's approach to aggregation, pertinent price-cost ratios are necessarily those that hypothetical monopolists would fix for those composites. That needs underlining, for in trying finally to delimit monopoly welfare losses Carson nowhere considers the problem thus posed. Rather in applying my (18) (his (3)) he apparently takes relevant monopoly price-cost ratios simply to be such as actually obtain for individual monopolized products. All things considered, perhaps the moral is that we must after all view Carson's results as representing an application of my (18) to individual products, rather than composites. From that standpoint, however, I wonder if he is right in holding that monopoly price-cost ratios generally are less than 1.20. So far as (1) applies, the relevant ϵ, by implication, must generally be greater than 6.0. Especially under oligopoly that seems high. But then again, if Carson's (3) is applied to individual products and λ_i is always taken (as he takes it) as given by $\epsilon \geq \eta_i$, oligopoly is indisputably not covered by Carson's results in any case.[3]

In the calculation of monopoly welfare losses, it has come to be a usual practice to refer only to consumer's goods. Explicitly or by implication, intermediate products are treated simply as if they were consumer's goods. In my article I fell in with this procedure, but pointed out that so far as there are intermediate products, the relevant λ_i for any consumer's good is not that established for that good. Rather it is that λ_i, together with corresponding margins on intermediate products used in the production of the consumer's good; or rather those margins to the extent that they contribute to the cost of the consumer's good. Carson refers to this fact (fn. 2) and notes that welfare losses might possibly rise by as much as 40 percent when account is taken of pricing distortions for intermediate products. As far as I can see, however, he nowhere considers such distortions in deriving his "maximum estimate." To the contrary, his calculations only become the more questionable when those pricing distortions are accounted for. Thus the λ_i that is relevant to monopoly welfare losses, that is, the one that takes account of pricing distortions for intermediate products, must generally be higher than (1) might indicate, if ϵ is still understood (as is proper), as the elasticity relevant to pricing decisions for one another consumer's good. Hence, the relevant λ_i exceeds the price-cost ratio indicated by (1) even under monopolistic competition and pure monopoly. Also, there is now a further reason for the pertinent λ_i to vary independently of pertinent η_i. For purposes of calculating losses, the latter still relate to consumer's goods rather than intermediate products.

III

In calculating monopoly welfare losses, Dean Worcester, like Carson, wishes to aggregate. While I have urged that reference is properly made to price-cost ratios and demand elasticities for individual products, there are, according to Worcester, theoretic grounds for referring instead, at least in the case of an elasticity, to the output of an "industry."

We must distinguish two kinds of welfare loss under monopoly. In one, the community realizes its production possibilities, but owing to monopolistic pricing practices produces a suboptimal output mix. In the other, at least in some industries, different firms produce a single, homogeneous product at differing marginal costs. As a result, the community fails to realize its production possibilities.

---

[3] I refer below to the data on industry elasticities compiled by Dean Worcester in his comment on my article. While such elasticities cannot generally be used, as Worcester would have us use them, in the calculation of monopoly welfare losses, they are suggestive of relevant magnitudes of ϵ under oligopoly.

In the real world, the second kind of waste can occur under competition, at least in the "short run," as well as under monopoly, and where it occurs under monopoly, it often coexists with waste of the first sort. Conceptually, however, the two kinds of waste are distinct, and in empirical work on monopoly welfare losses, attention has usually been focused on waste of only the first sort, that associated with a suboptimal output mix. That is apparently the concern in all the calculations that I sought to evaluate in my original article. In trying to provide an analytic framework for the appraisal of such calculations, therefore, I simply assumed that the community is in fact on its production possibilities schedule. Only waste associated with a suboptimal output mix was considered.

As may not have been stressed sufficiently, however, from that standpoint aggregation to a certain point is not only not barred, but indicated. Thus, for optimality of the output mix, where different firms produce identical products necessarily only the total output of all firms together can matter. True, how that output is divided among the different firms may itself be a source of waste under monopoly, but that is the kind of waste where production possibilities are not realized rather than that where the output mix is suboptimal. It also follows that where reference is to optimality of the output mix and different firms produce identical products, it is the elasticity of demand for their total output that must be considered in calculating relevant losses.

All this must be kept in mind in considering Worcester's theoretic grounds for urging reference to industry rather than product demand elasticities in the calculation of monopoly welfare losses. Those grounds revolve around his analysis of two theoretic models of oligopoly. One of these, the "independent pricing" model, is of the familiar textbook sort where a dominant firm fixes prices monopolistically in the light of the aggregate demand for the output of the industry after making due allowance for supplies forthcoming from other firms that in effect act as perfect competitors. Although Worcester is not explicit, he clearly assumes here, as is often done, that the different firms are producing identical products. That is evident from his analysis of price determination in the industry, and also from his apparent supposition that, costs apart, it is immaterial for welfare how output is divided among the different firms.

It follows that at this point there can be no difference between Worcester and me on the principal question at issue. He contends and I of course agree that in the calculation of monopoly welfare losses, it is the elasticity of demand for the output of the industry that must be considered, so far as different firms produce identical products. Or rather I agree to the extent that the concern is with losses associated with a suboptimal output mix. In the case which Worcester considers, there evidently is also a shortfall from production possibilities, and Worcester might be read to mean that waste of both sorts together turns simply on the elasticity of demand for the industry's output. While a shortfall from production possibilities was not of concern in my original article, it should be observed that under dominant firm pricing such as in question, total waste properly viewed depends on not only the elasticity of demand for the industry's product, but the comparative cost schedules of different firms.

Worcester also considers a "collusive" model of oligopoly, where "each firm immediately matches the price of any other firm in the group, perhaps that set by a price leader." Under this model, we are told, "industry elasticities . . . are identical to firm elasticities and therefore are appropriate to the calculation of welfare loss. . . . " Here again Worcester is not entirely explicit regarding the comparative nature of the products of different firms, but one might wonder whether the collusive matching of prices that is envisaged does not already presuppose that those products are identical. In any event, his conclusion that it is the industry elasticity that must be considered in calculating losses is surely a nonsequitur on any other assumption.[4] Actually, Worcester himself apparently recognizes that his conclusion does not follow so far as there is "nonprice competition." That presumably prevails so far as there is product variety of any sort of interest to consumers. Hence, here too, I find no basis to disagree with him.

What, then, is in dispute? I am not sure, but Worcester may be assuming that oligopolistic firms characteristically produce identical products. That can hardly be so. As indicated, I understand oligopoly to prevail where a relatively limited number of firms produce more or less similar, though not necessarily identical, products. In fixing prices, at least some of the firms consider the impact of their actions on rivals. That I believe is a rather conventional conception. On that understanding, the products of different firms in an oligopolistic industry may be identical, but they are certainly not characteristically so. One may wonder whether they are typically so in oligopolistic industries generally, but in the calculation of monopoly welfare losses it is demand elasticities for consumer's goods that are of immediate concern. In respect of such goods, product variety is if anything virtually a hallmark of oligopolistic industries.

Possibly Worcester assumes that differences in products prevailing in such industries are immaterial to welfare, and that that is so even though consumers themselves value those differences. Such a standpoint is not exactly a rarity today, but it represents a major breach in the principle of consumers' sovereignty. That principle has always been presupposed hitherto in calculations of monopoly welfare losses.

While Worcester focuses on oligopoly, he contends that industry elasticities are the ones to refer to in the calculation of losses under monopoly generally. Apparently, that is held to be appropriate even under monopolistic competition understood in the usual sense that I explained above. I find Worcester very puzzling at this point, for he nowhere provides any support for such a view, and indeed sometimes seems to recognize its error. Needless to say, there is no basis here for me to shift from the standpoint I have taken on the question at issue.

Worcester has put us in his debt, then, by compiling, as he has done, systematic data on industry demand elasticities, but it is not possible to follow him in supposing they are applicable to the calculation of monopoly welfare losses. It also follows that his industry elasticities cannot readily be used, as Worcester uses them, as a benchmark for judging the plausibility of the hypothetical elasticities that I considered in making the illustrative calculations in my paper. Those elasticities are intended to relate to products, not industries.

In the light of diverse information that he also compiles on price-cost ratios, Worcester concludes that "We leave the matter of size of monopoly markups unresolved, but with considerable skepticism attached to values which require net income above competitive levels to exceed 15 percent of sales for more than a few years." I agree that the matter has been left unresolved.

While Worcester questions my hypothetical elasticities, a "much more serious objection" to my illustrative calculations is held to be that "most of the combinations of monopoly markup and elasticity" that I consider "are entirely inconsistent with profit-maximizing behavior." Apparently reference essentially is to the fact that I allow combinations of λ_i and η_i that fail to conform to (1) when ϵ is taken to equal η_i. As indicated in my reply to Carson, however, such a departure from (1) seems warranted where oligopoly prevails, for there, if (1) applies at all, ϵ is apt to differ from η_i, and sometimes markedly. Or rather, as we must now be careful to note, that is so except where oligopolistic firms produce identical products, which is the case on which Worcester implicitly focuses. Departures from (1) with $\epsilon=\eta_i$ are the more in order so far as η_i relates to consumers goods while λ_i properly reflects monopoly profit margins on intermediate products as well. Worcester neglects that aspect throughout.

---

[4] Should price-cost ratios of all firms be the same, however, and production be under conditions of constant costs, the output of the industry might be treated as a single composite even if the firms produce different products. See the author, p. 860.

Worcester holds that I even contemplate some equilibria that entail negative outputs of some products, or, what comes to the same thing, are beyond the community's production possibilities. The difficulty with Worcester's numerical demonstration of that odd result is that he erroneously takes a point elasticity of demand to apply over a wide range. If he will redo his calculations to correct that error, I think he will see that the community in fact remains on its production possibilities and so produces only positive outputs throughout, though under his assumptions no doubt it is pushed virtually to an extremity. More generally, all of my illustrative calculations were made to allow the community to realize its production possibilities both with and without monopolistic pricing in monopoly sectors.

IV

What follows regarding monopoly welfare losses? In presenting illustrative calculations on that matter in my original article, I of course did not mean to suggest that the more extreme cases considered were apt to be typical. With the information at hand, it simply did not seem wise to try to restrict narrowly the admissible possibilities. Carson's "maximum estimate" for monopoly welfare losses, however, is 3.2 percent of the national income. Even if that figure were accepted as an upper limit for the losses in question, it is worth noting that they could then be as much as thirty-two times as large as Harberger and Schwartzman found in calculations that I question. Such losses would also amount today to $45 billions. But, despite Carson's ingenious computations, it is difficult to avoid the conclusion that an upper limit for monopoly welfare losses still remains to be fixed. It does not seem possible either to join Worcester in thinking that "The time is surely close at hand for economists to accept the research findings that welfare losses due to enterprise monopoly are small in the American economy. On the evidence, industry approximates the competitive model. . . . "

In sum, a principal moral of my 1973 paper was clearly that, contrary to Harberger, Schwartzman, and others, monopoly welfare losses could easily be quite consequential. I also urged, however that more research is needed on the important question at issue. Illuminating as the comments of Carson and Worcester are, I must conclude that both morals are still indicated.

REFERENCES

A. Bergson, "On Monopoly Welfare Losses," *Amer. Econ. Rev.*, Dec. 1973, *63*, 853–70.

R. Carson, "On Monopoly Welfare Losses: Comment," *Amer. Econ. Rev.*, Dec. 1975, *65*, 1008–14.

A. C. Harberger, "Monopoly and Resource Allocation," *Amer. Econ. Rev. Proc.*, May 1954, *44*, 77–87.

D. R. Kamerschen, "Welfare Losses from Monopoly," *Western Econ. J.*, Summer 1964, *4*, 221–36.

D. A. Worcester, Jr., "On Monopoly Welfare Losses: Comment." Dec. 1975, *65*, 1015-23.

The Economics of the Network-Affiliate Relationship: Comment

By DANIEL A. GRAHAM AND JOHN M. VERNON*

In a recent issue of this *Review*, Stanley Besen and Ronald Soligo (hereafter, B-S) presented an ingenius graphical analysis of the economics of network-affiliate relationship in the television broadcasting industry. The purpose of this note is to correct several errors in their analysis, to "adjust" their basic diagram in a way that makes certain results much more transparent, and to demonstrate that the relaxation of one assumption invalidates one of their major conclusions.

I. The Besen-Soligo Model

We begin by describing the analysis of B-S and our extension of it. The focal point is the relationship between a network and a "representative" affiliate for the case in which the number of local stations exceeds the number of networks. In this context B-S assume that the network has monopsony power in the purchase of *cleared time* from the affiliate. The affiliate is viewed as supplying a service to the network, namely, the broadcasting of network programs, and the total supply of this service is referred to as cleared time. Affiliates are compensated in two ways: 1) by receiving a share of the revenue that the network collects from its sale of advertising time on network programs, and 2) by receiving a share of the total time available for commercials within each network program. (The affiliate retains all revenue from its sale of this commercial "spot time.")

For the ith hour of programming, the affiliate can calculate that value of p (the fraction of network advertising revenues returned to the affiliate) for which it would be indifferent between clearing the time to the network and using the time for local programming. This break-even value of p is labeled p_i by B-S and is given by the condition:

$$p_iA_i + S_i = V_i \tag{1}$$

where

$A_i \equiv$ network advertising revenues for the ith hour[1]

$S_i \equiv$ affiliate advertising revenues from the sale of spot time during the ith hour

$V_i \equiv$ net revenue that could be earned by the affiliate by broadcasting local programming during the ith hour.

The B-S analysis may be clarified somewhat if we assume that the network and affiliate can sell commercial time on a given program for the same price[2] and if we let β_i denote the ratio of spot time to network advertising time during the ith hour. We may then write

$$A_i = T_i/(1+\beta_i)$$
$$S_i = \beta_iT_i/(1+\beta_i)$$

where $T_i =$ the total value of advertising time on the ith hour of network programming. If we further assume, as in fact appears to be the case,[3] that β_i does not vary from hour to hour, then equation (1) may be rewritten as

$$p_i = (1+\beta)V_i/T_i - \beta \tag{2}$$

where β denotes the common value of the β_i.

The heart of the B-S analysis is the affiliate's supply curve of cleared time. To construct this curve, array program hours in order of increasing p_i or, equivalently, in order of increasing V_i/T_i. The supply curve is then obtained as the curve described by the coordinate pairs $(\sum_{j=1}^{i} A_j,\ \hat{p}_i)$. This supply curve relates the quantity of cleared time supplied (measured in terms of the associated value of network advertising) to the revenue-share p set by the network.

---

* Associate professor and professor of economics, respectively, Duke University. We are indebted to Merton Peck for helpful comments on an earlier draft.

[1] This is not total network advertising receipts but only that portion attributable to the region serviced by the particular affiliate.

[2] Notice that both affiliate and network are selling what is inherently the same commodity, e.g., a "New York, Gunsmoke, viewer minute."

[3] See Roger Noll, Merton Peck, and John McGowan, p. 61.

Such a supply curve is illustrated by the curve *BC* in Figure 1. If the network should set p_0, for example, the affiliate would clear all hours to the left of A_0 since p_0 exceeds the break-even $\hat{p}_i$ for each of these hours. In such circumstances B-S point out that the network's share of advertising revenues would be area *II* and the affiliate's share would be the sum of areas I_a and I_b.

They also point out that the net profit or quasi rent to the station from network affiliation is given by the sum of areas I_a and *III*. Although this last observation is correct, it can be made much more transparent by adding a horizontal line to their diagram at a distance β below the horizontal axis. Regarding this line βW as the new horizontal axis, we may interpret the area under the supply curve and to the left of A_0 as the opportunity cost to the affiliate of supplying A_0 of cleared time. This can be seen as follows. The ith hour accounts for network advertising revenues equal to A_i or, in Figure 1, for that portion of the horizontal axis corresponding to the width of rectangle *FR*. Since the height of this rectangle from *T* to *R* is equal to β or S_i/A_i and since the width is equal to A_i, the area from *T* to *R* must be equal to S_i. Recalling equation (1), to break even on this hour $\hat{p}_i A_i$ (the area of the rectangle from *J* to *T*) plus S_i (the area from *T* to *R*) must be equal to V_i. Thus, by construction, the area from *J* to *R* must be V_i, the opportunity cost to the affiliate of clearing the ith hour. Summing over the total amount of cleared time A_0, we obtain the area under *BC* and above the new axis βW as the total opportunity cost. It should now be clear that the total revenue received by the affiliate from its share of network advertising and from its sale of spot time (the area of rectangle $\beta p_0 MW$) exceeds the total opportunity cost by the sum of areas I_a and *III* (the affiliate's quasi rent).

The addition of the βW line would have enabled B-S to avoid an error of interpretation. On page 263 they refer to the area under the supply curve and above ***their*** axis OA_0 (comparable to area I_b in Figure 1) as the profits on hours to the right of *X* that "the station would earn if it were an independent station." This is clearly in error since it understates the true opportunity cost by an amount equal to the value of the spot time for these hours.

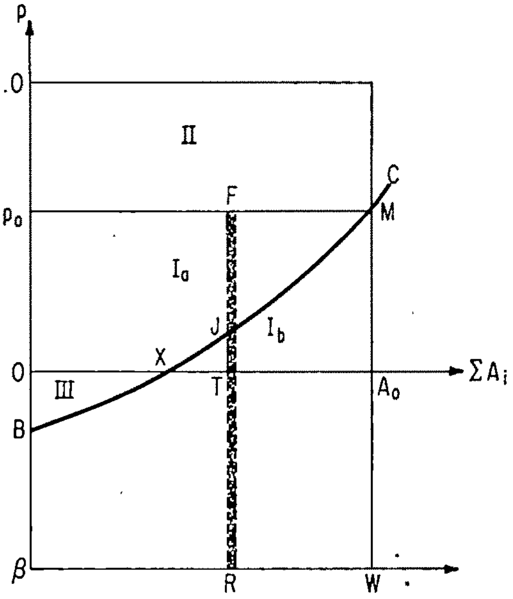


FIGURE 1

II. The Nondiscriminating Monopsonist Solution

The B-S solution to the network profit-maximization problem is reproduced in Figure 2. They assume that the marginal cost of supplying an additional program to the affiliate is zero. Thus profits are maximized when the curve marginal to the supply curve intersects the horizontal line at $p=1$. This point is characterized by the fact that the marginal revenue product to the network of the last program hour is equal to the zero marginal cost of programming. The network maximizes profits, therefore, by setting a revenue-share price equal to p_e and by purchasing A_e of cleared time from the affiliate.

There appear to be two errors in this

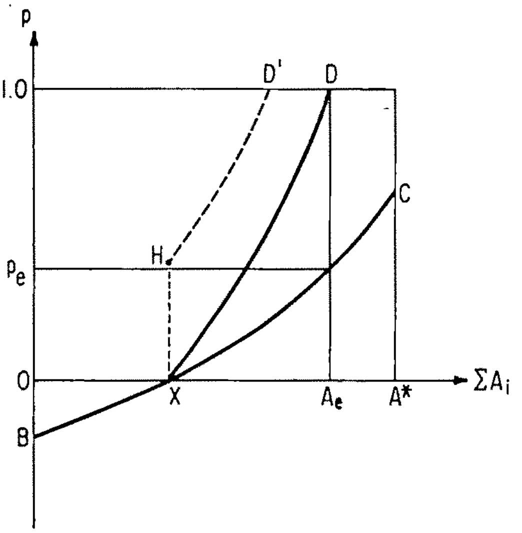


FIGURE 2

analysis. First, B-S argue that the curve "marginal to" BC is given by XD:

> The curve XD does not extend below the horizontal axis because the network will have programs of value OX cleared even if $p=0$. If the network were permitted to make p negative, i.e., to charge the affiliate for a cleared program, the curve marginal to the supply curve would begin at B. [p. 262, fn. 8]

The fact that negative p values are not allowed, however, has nothing to do with where the marginal curve begins. The correct marginal curve is essentially the same as the marginal outlay curve of a monopsonist confronting a factor supply curve subject to a minimum factor price constraint. In Figure 2 the correct curve coincides with the horizontal axis from O to X, skips discontinuously from X to H, and coincides with the dashed curve HD' thereafter. If negative p were allowed, HD' would extend leftward to point B. This, of course, does not alter the fact that the correct marginal curve must "pass through" point H in either circumstance.

The second deficiency in the B-S analysis involves the assumption (pp. 260, 262) that all network programming costs are fixed and that the marginal cost of supplying an additional hour of existing programming is therefore zero. There are at least two problems with the assumption of zero marginal programming costs.

First, according to B-S, Figure 2 depicts a *representative* affiliate. Presumably, therefore, we can regard *all* affiliates as reaching an equilibrium relationship with the network characterized by the intersection of the marginal curve and the horizontal line at $p=1$ (where marginal programming costs are zero). But this implies that the network *overinvested* in programming. That is, if the marginal revenue product to the network of the last hour cleared by a representative affiliate (and thus by *all* affiliates) is to be zero, then the network would have been better off had it not invested in that last hour of programming.

The second problem with the assumption of zero marginal programming costs is that it is conceptually misleading. Zero marginal programming costs imply fixed programming costs during the period in which the network sets its pricing policy concerning affiliates. However, at repeated intervals the network must enter into agreements to purchase programs. At such times the network must answer questions of the sort, "How much is program ______ worth?" The answers to such questions are, of course, dependent upon the particular pricing policy that the network will be able to employ vis-à-vis its affiliates. The network must, therefore, determine its optimal pricing policy toward affiliates interdependently with the determination of its demand for programs.

We would argue, therefore, that the actual solution of the network pricing problem would be made at the time that programs are purchased and would involve setting a revenue sharing price for affiliates such that the marginal revenue product to the network of the last hour is equal to the nonzero cost of purchasing that hour of programs. Since, as Noll, Peck, and McGowen indicate (p. 66), the cost of programs may amount to 30 percent of total network advertising revenue, the difference between this solution and that of B-S may be substantial.

One further aspect of the nondiscriminating monopsonist solution should be mentioned. While B-S admit that the proportion (β) of commercial time given to an affiliate is an important dimension of the affiliation agreement (p. 259), they assume, presumably for purposes of simplicity, that this proportion is fixed and not subject to change by the network (p. 260). While, as we shall see, this assumption has no appreciable effect upon the nondiscriminating monopsonist solution, it does have a rather significant effect if the network can discriminate. Turning first to the nondiscriminating case, we observe that equation (2) can be rewritten as

$$\frac{p_i+\beta}{1+\beta}=\frac{V_i}{T_i}$$

This means that if the network were to set any p, β combination satisfying

$$\frac{p+\beta}{1+\beta}=\frac{V_{i'}}{T_{i'}}$$

a total of i' hours of cleared time would be supplied by the affiliate. Thus $\gamma\equiv(p+\beta)/(1+\beta)$ can be regarded as the parameter which determines the supply of cleared time and the supply function can be written

$$t=f(\gamma)$$

where $t\equiv$supply of cleared time (demand for network programming) by affiliate measured in hours. The network profit maximization problem thus becomes

$$\max_{p,\beta}\pi\equiv\sum_{i=1}^{f(\gamma)}[(1-p)A_i-C_i]$$

where $C_i\equiv$the cost to the network of obtaining the ith hour of programming.[4]

Substituting for A_i yields

$$\pi=\sum_{i=1}^{f(\gamma)}[T_i(1-p)/(1+\beta)-C_i]$$
$$=\sum_{i=1}^{f(\gamma)}[T_i(1-\gamma)-C_i]$$

The network problem of selecting p and β so as to maximize profit thus reduces to one of selecting γ. Denoting this profit maximizing γ by γ^* it follows that the network would be indifferent among all p, β combinations which satisfy

$$\frac{p+\beta}{1+\beta}=\gamma^* \tag{3}$$

Since all such pricing policies have already been shown to entail the same supply of cleared time, the ability to set β adds nothing of interest to the solution. Given an arbitrary β the network would do just as well by simply solving equation (3) for the associated p; the total supply of cleared time, network profit, and affiliate profit would remain invariant under all such solutions.

III. The Discriminating Monopsonist Solution

B-S show that the ability of the network to discriminate, i.e., to extract portions of the quasi rents that accrue to affiliates, hinges upon the use of two devices: 1) *uncompensated time*, a two-part pricing scheme in which the network sets a zero price for an initial number of hours cleared and a positive price for all subsequent hours; and 2) *option time*, a scheme in which affiliates are required to clear a certain minimum number of hours in order to maintain their affiliation.

Using graphical analysis they argue that while the use of uncompensated time allows the network to capture a portion of the quasi rent, it is necessary for the network also to use option time if it wishes to extract all rents. Referring to the Federal Communications Commission ban on the use of option time B-S conclude: "The bar against option time, if it is effective, may result in some quasi rents being retained by the affiliate" (p. 265). As support for this conclusion, they go on to cite the empirical finding that network affiliates are more profitable than independents when other determinants of profitability are taken into account.

The B-S argument can be seen by reference to Figure 3. Here point $\bar{A}$ is chosen so that area I is equal to area II. Suppose now that the network wants to buy A^* of cleared

---

[4] Here we assume that all n affiliates are alike. Thus C_i, the program cost for one affiliate, can simply be regarded as one nth of the total program cost.

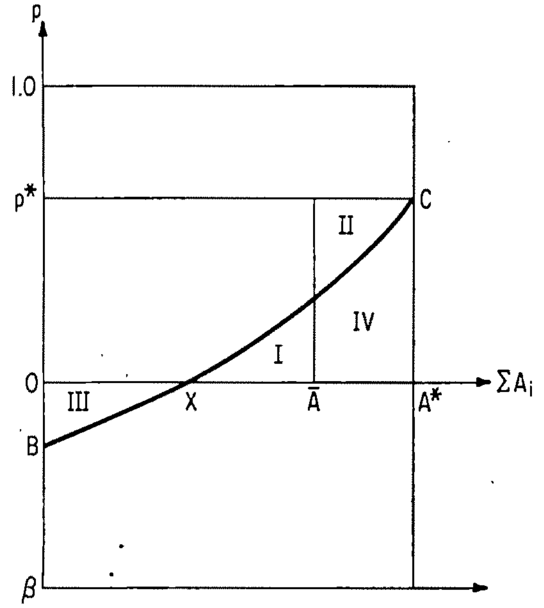


FIGURE 3

time from the affiliate. By setting p equal to zero for the first $\bar{A}$ of cleared time and by setting p equal to p^* for all subsequent hours the network would render the affiliate indifferent between clearing X hours and A^* hours. Thus the network would capture the quasi rents on all hours to the right of X. As B-S point out, however, quasi rents equal to area *III* still accrue to the affiliate. The network, moreover, is apparently unable to capture this remaining rent since increasing the amount of uncompensated time would induce the affiliate to clear only X hours and since reducing p below p^* on subsequent hours would also induce the affiliate to clear less than A^* hours. These considerations are the basis for the B-S conclusion that the use of option time is necessary to capture the remaining quasi rent.

We find this conclusion to be unwarranted, however, since it depends crucially upon the simplifying assumption that the network cannot control β. If the network can in fact determine β then it can extract all the quasi rent with the use of uncompensated time alone. This is done by reducing the value of β. Such reductions in β shift the affiliate supply curve in such a way that point X moves leftward. By continuing until point X coincides with the origin, the network can eliminate area *III* entirely. It can then determine $\bar{A}$ so that quasi rents on *all* hours are captured by the network. More precisely, by setting

$$\beta^* = \frac{V_1}{T_1 - V_1} > 0$$

it follows from equation (2) that

$$p_1 = (1 + \beta^*)\frac{V_1}{T_1} - \beta^* = 0$$

This choice of β thus produces an affiliate supply curve which intersects the horizontal axis at the origin. Now suppose that the network wishes to clear t^* of programs measured in hours. This requires that the network set

$$p^* = (1 + \beta^*)\frac{V_{t^*}}{T_{t^*}} - \beta^*$$

By then selecting that number of uncompensated hours which equates the quasi rent on subsequent hours to the losses on the uncompensated hours, the network can capture all quasi rent without recourse to option time.

REFERENCES

S. M. Besen and R. Soligo, "The Economics of the Network-Affiliate Relationship in the Television Broadcasting Industry," *Amer. Econ. Rev.*, June 1973, *63*, 259–68.

R. G. Noll, M. J. Peck, and J. J. McGowan, *Economic Aspects of Television Regulation*, Washington 1973.

The Economics of the Network-Affiliate Relationship: Reply

By STANLEY M. BESEN AND RONALD SOLIGO*

Daniel Graham and John Vernon's (G-V) comment on our paper raises four points: 1) that we have erred in the construction of the curve marginal to our "supply" of time schedule; 2) that we have mistakenly identified the area under the supply curve as the return which a network affiliate would have received had it remained an independent station; 3) that if the network regards the marginal cost of supplying a program to an additional affiliate as zero that this will lead to an oversupply of programming; and 4) that when our analysis is extended to permit variations in the distribution of advertising time between the network and the affiliate that one of our conclusions, that the ban on option time permits the affiliate to earn above normal profits, is vitiated.

We are grateful to G-V for pointing out the technical slip in the construction of the curve marginal to the supply schedule and for indicating the inaccuracy in our description of area under the supply curve. We would point out only that while our description was inaccurate, our analysis was correct.

G-V's third point is not correct. They are clearly right that the network must determine whether it should produce a program at all. And, clearly, for the network to be willing to do so the sum of the returns it gets from each of the stations that carry the program (the network share of advertising revenue minus the payment to the affiliate) must exceed the cost of carrying the program. But, contrary to their claim, the network need not have "*overinvested* in programming" even " . . . if the marginal revenue product to the network of the last hour cleared by a representative affiliate (and thus by *all* affiliates) is . . . zero. . . . " (p. 1034).

To see why this is so it is necessary only to recognize that the array of programs according to the sharing ratio which will induce the affiliate to accept programs can and in general will be different for different stations. The relative attractiveness of a given network program and its nonnetwork programming alternatives is likely to differ from market to market because, for example, audiences tastes differ. If these differences exist, the return to the network from the carriage of the marginal program by each affiliate *can* be zero while all programs in the network schedule are profitable. A program which is marginal for one station will be inframarginal for others and there will be contributions to network profits from the carriage of the program by all stations for which the program is inframarginal.

G-V's analysis of this point is misleading in one important respect. It is generally not possible to write down an equation such as their profit-maximizing equation since it is not possible to know on the basis of data from a single station alone what share of the cost of a program it should bear if the network is to maximize its profits. For the case cited in G-V's footnote 4 in which all affiliates are identical, equal apportionment of cost gives the correct answer, but in general it will not be possible to employ this procedure.

While G-V are correct that the network must earn positive profits on each program, their attempt to take production costs into account at the time of establishing the terms of the affiliation contract runs the risk that the network will not produce a program which would increase its profits. The approach that we suggest, that the network first decide what its payment will be to each affiliate *if a program is produced* and then decide whether the program should be produced at all, always yields the correct answer

---

* Professors of economics, Rice University. We wish to acknowledge financial support from National Science Foundation, grant GI 38909.

while the proposed G-V procedure does not.

As for point 4), G-V are correct that the network can, by reducing the proportion of advertising time during which an affiliate can sell advertising, shift the supply curve of programming upward, thus increasing network profits while leaving the amount of programming accepted unchanged. This represents an interesting extension of our model. We would, however, raise two points concerning it. First, the policy requires that the networks increase the proportion of network advertising revenue p paid to affiliates to get them to clear the entire network schedule.[1] But, as we pointed out in our paper (p. 264), there seems to exist an informal upper bound on p which is designed to make it possible for nonnetwork program suppliers to have access to the time of network affiliates. If the increase in p required by the reduction in station advertising time would cause it to exceed this limit, it may still prove impossible for the network to get all of a station's quasi rents. Second, since presumably one of the objectives of the ban on option time was to increase station profits, the Federal Communications Commission might not be terribly receptive to changes in network behavior which offset the effects of the ban, especially when the objective of the change is likely to be transparent. The basic thrust of G-V's point 4) deserves emphasis, however. When the regulated firm has many degrees of freedom, it will often be able to partially or totally frustrate the efforts of the regulators.

REFERENCES

S. M. Besen and R. Soligo, "The Economics of the Network-Affiliate Relationship in the Television Broadcasting Industry," *Amer. Econ. Rev.*, June 1973, *63*, 259–68.

D. A. Graham and J. M. Vernon, "The Economics of the Network-Affiliate Relationship: Comment," *Amer. Econ. Rev.*, Dec. 1975, *65*, 1032–36.

---

[1] Network profits increase, nevertheless, because the *total* amount of network advertising revenue increases.

NOTES

On March 14, 1975 the Executive Committee of the American Economic Association approved the following changes in the financial operation of the *American Economic Review* and *Journal of Economic Literature.*

a) The discontinuance of free reprints/offprints to authors of material appearing in the *American Economic Review* and *Journal of Economic Literature.* Reprints/offprints will continue to be sold at prices which cover the relevant costs. This change will begin with the December 1975 issues.

b) A differential submission fee for authors who are not members of the American Economic Association. As of July 1, 1975, the submission fee for manuscripts submitted to the *American Economic Review* is now fifteen dollars ($15) for members and thirty dollars ($30) for nonmembers.

c) A postage surcharge for mailing of material outside the United States. This surcharge will equal the difference between mailing costs inside and outside the United States. It will be added to new subscription and membership bills January 1, 1976.

d) A page charge for articles, shorter papers, notes, comments, and replies published in the *American Economic Review* and in the *Journal of Economic Literature.* The charge will be twenty-five dollars ($25) per printed page, and is in effect with manuscripts accepted after July 1, 1975.

The page charge is payable by the institution or granting agency supporting the research. Payment of this charge is not a prerequisite for publication, nor are authors expected to pay these charges themselves. Moreover, articles are accepted without prior knowledge of any financial support which the author may have at his disposal. If the author or his institution cannot pay the page charge, it is no obstacle to publication. The purpose of the page charge is to take advantage of willingness of sponsoring agencies in the federal and state governments and private foundations to defray the expense of publishing research in scholarly journals. Page charges are regarded as a legitimate item in the budgets of research grants, and journals in other disciplines have taken advantage of this source of revenue for some time.

Annual Meeting Placement Service

The Placement Service at the 1975 annual meetings of the Allied Social Science Associations in Dallas will begin operation on December 27, the day before sessions begin. Applicants and employers will be able to attend more sessions with a day set aside entirely for labor market transactions. The Placement Service will continue during the other days of the annual meetings, but hopefully at a less frantic pace than in previous years. This service will be located at the Convention Placement Center in the Dallas Convention Center. It will be open from 10:00 A.M. to 5:00 P.M., December 27; 8:30 A.M. to 5:00 P.M., December 28–29; and 8:30 A.M. to 12:00 noon December 30.

For your convenience the National Registry for Economists is offering a pre-convention registration service. Forwarding your application or position(s) to the Registry prior to the meetings will expedite service at the Convention Placement Center. Employment applications and listings of available positions will also be accepted from those who cannot attend the meetings and will be made available for review by the participants. Immediately upon arrival at the convention, registrants should report to the Convention Placement Center so that their pre-convention registration can be activated. Forms may be obtained from the National Registry for Economists, Illinois State Employment Service, 40 West Adams Street, Chicago, Illinois 60603. They must be completed and returned to the National Registry prior to December 5, 1975.

The program committee for the September 1976 meetings of the American Economic Association invites contributed papers. These will be screened by committee members and notification of acceptance or rejection will be sent by March 1, 1976. Prospective contributors are urged not to submit similar papers to the American Economic Association, the Econometric Society, or other associations. Contributed papers for the American Economic Association sessions should be of the type that would not be considered primarily for presentation to the Econometric Society. Those interested in submitting papers for consideration are invited to send abstracts *not later than* February 1 to Lawrence R. Klein, Department of Economics, University of Pennsylvania, Philadelphia, Pennsylvania 19104.

Economists who are *strongly* oriented toward the humanities, who use humanistic methods in their research, and who will be participating in meetings abroad that are concerned with the humanistic aspects of their discipline are eligible to apply for travel grants of the American Council of Learned Societies. Specifically, economists may be eligible if (a) they deal with the history of economic thought or economic history, and (b) if their approach is qualitative and descriptive rather than quantitative and statistical. Conferences dealing with the establishment of social policy or legislation are generally ineligible. The deadlines for applications to be received in the office of the American Economic Association are: for meetings scheduled between June and September, February 15; for meetings scheduled between October and January, June 15; for meetings scheduled between February and May, October 15. Application forms may be obtained from Rendigs Fels, Secretary, American Economic Association, 1313 21st Avenue South, Nashville, Tennessee 37212.

The Joint Committee on Soviet Studies and the Joint Committee on Eastern Europe of the American Council of Learned Societies and the Social Science Research Council wish to draw attention to Grants Programs in Soviet Studies and East European Studies. The programs include Grants for Post-Doctoral Research, Grants for Study of East European Languages, Grants in Support of Conferences, and Travel Grants to International Conferences Abroad. For details of eligibility and information which *must* be supplied in requesting application forms, request an announcement brochure from Office of Fellowships and Grants, American Council of Learned Societies, 345 E. 46th Street, New York, NY 10017.

As of July 1, 1975, the *Journal of International Business Studies*, the official journal of the Academy of International Business, is being edited and published at the Rutgers School of Business. All persons engaged in the preparation of high quality advanced scholarly research papers in the areas of international business, management, institutions, economics, and policy are invited to submit manuscripts in three copies for editorial consideration and possible publication to Professor Robert Vambery, Rutgers University, Graduate School of Business, Newark, NJ 07102.

Omicron Delta Epsilon, the International Honor Society in Economics, invites the submission of entries for the eighth year of the Irving Fisher Graduate Monograph and Frank W. Taussig Award Undergraduate Competitions. The Fisher Award consists of $1,000 and publication as a book by Princeton University Press, subject to approval of its editorial board. In addition, the winner will be invited to submit a paper based on the winning entry to the *American Economic Review*. The recommendations of the Final Selection Board of the Competition will be considered by the *Review* in the refereeing process. All finalists will be invited to submit a paper for publication in *The American Economist*. The Taussig Award consists of $100 and publication in *The American Economist*. Entries for the Fisher Award should be submitted to Departmental Selection Committees by January 1, 1976 and entries for the Taussig Award by May 15, 1976. They will be judged by the International Editorial Board and finalists by the Final Selection Board, consisting of Professors William J. Baumol, Frank H. Hahn, Robert M. Solow, Arnold Zellner, and Egon Neuberger (editor). For more information, write Professor E. Neuberger, Editor, Economic Research Bureau, State University of New York, Stony Brook, NY 11794.

Nijenrode Congress: Call for Papers

From August 25 to 27, 1976, a Congress on Decision Making in Business will be held at The Netherlands School of Business in Breukelen, The Netherlands. Papers are invited in the various subdivisions of management studies: accounting, finance, marketing, organization, law, environment, ethics, psychology, sociology, communication, economics, O.R., etc. The Congress will consist of two sessions: a one-day symposium, during which prominent speakers will address and take part in discussions with a large gathering of participants; and a two-day research seminar, consisting of parallel sessions devoted to particular themes or topics. Those who wish to present papers are invited to send a resume to Dr. C. van Dam, The Netherlands School of Business, Nijenrode, Breukelen, The Netherlands. Preliminary drafts of papers will be needed in triplicate before February 1, 1976.

The S. S. Huebner Foundation for Insurance Education is initiating an experimental program of grants to support scholarly research in the field of risk and insurance. Full-time faculty at colleges and universities in the United States and Canada are eligible to apply for grants. Applicants must hold a doctoral degree or be a fellow of an actuarial society. Proposals will be considered for projects in all areas relating to insurance including, but not limited to, the role of insurance companies in the capital markets, the theory of consumer demand for insurance, and social security and other governmental insurance programs. Awards of up to $10,000 are available, and grant applications are due by February 1, 1976. For further information contact Dr. J. David Cummins, Research Director, S. S. Huebner Foundation for Insurance Education, W-133 Dietrich Hall/CC, University of Pennsylvania, Philadelphia, Pa. 19174.

Call for Papers

The XIV International Symposium on the Application of Computer Methods in the Mineral Industries (*APCOM*) will be held at the Pennsylvania State University October 4–8, 1976. The program committee invites abstracts of papers to be considered for presentation at the symposium. The abstract must not be more than 500 words, and should be mailed to reach Dr. R. V. Ramani, 116 Mineral Industries Building, University Park, Pa., 16802, U.S.A., not later than January 15, 1976. Sessions on the following topics, as they pertain to the mineral industries, are being planned: national and international planning for raw materials; investment planning in multinational companies; health, safety, environment and manpower planning; computers in coal and mineral processing; exploration and ore reserve estimation; production planning and control; management information systems; industrial case studies; automation and process control; economic modeling of the mineral sector. Those interested in receiving future mailings should send their names and addresses to Dr. R. V. Ramani.

The Social Science Research Council announces two postdoctoral fellowships in criminal justice indicators for the academic year 1976–77. The program provides an opportunity for fellows to work on analysis and indicator development projects of their own design, using statistical data from the Law Enforcement Assistance Administration or other sources. In selecting fellows, the subcommittee will give priority to projects emphasizing the measurement of change and the development and analysis of time-series, particularly series based on replicated surveys. There are no application forms. Applicants are asked to submit resumes, copies of publications, and project proposals (indicating the questions to be investigated, the data base to be used, and the plan of analysis) no later than January 31, 1976, to David Seidman, SSRC Center for Social Indicators, 1755 Mass. Ave., N.W., Washington, D.C. 20036 (202+667-8884).

April 21 through 24, the Society of Government Economists and the Economic Development Administration of the Commonwealth of Puerto Rico will hold a conference on economic development in San Juan, Puerto Rico. Sessions on financial incentives and economic development techniques will focus on problems relating to transportation, health, energy, education, and labor. Sessions are also planned on data requirements, economic models, socio-cultural barriers to development and research. For program information write to Tom Herrick, Long Island University, Brooklyn, New York, or Lewis Smith, Director of Planning and Research, Economic Development Administration, San Juan, Puerto Rico. Registration information can be obtained by writing Martin Stein of the Maryland Department of Transportation. P.O. Box 8755, Baltimore-Washington International Airport, Baltimore, Maryland 21240.

The fourth national conference sponsored by the Intersociety Committee on Transportation will be held in Los Angeles, California, July 18–23, 1976. Papers will be presented covering all aspects and modes of ground, air, and water transportation. Planned sessions include air cargo, highway and rail freight, passenger modes, water transportation, mass transit, pipelines, multimodal, transportation of energy, economics and costs, policy and planning, command and control, environmental impact, energy considerations, information systems, design and engineering, operations. Authors are invited to send abstracts of 300 words or more on these or other related areas by December 15, 1975 to J. F. Courtney, Vought Systems Division, P.O. Box 5907, Dallas, Texas 75222.

A Committee on the Social Sciences in the National Science Foundation has been appointed by the National Academy of Sciences in response to a request from the Director of the National Science Foundation. The committee has been asked to submit a report to NSF that includes recommendations regarding research areas and activities that should be supported in the future. The committee invites members and readers to comment on and offer suggestions for promising directions of social science research and activities that NSF should consider strengthening or initiating. Since the committee will begin to draft its report late this year, please write within the next few weeks to Herbert A. Simon, Chairman, National Academy of Sciences, JH 818, 2101 Constitution Avenue, Washington, D.C. 20418.

Deaths

Harry Gunnison Brown, professor emeritus of economics, University of Missouri, former acting dean; and past president of the Midwest Economic Association, Mar. 18, 1975.

Barbara Carbaugh, Columbus, Georgia, May 17, 1975.

Thor Hultgren, department of economics, University of Wisconsin-Milwaukee, July 31, 1975.

Retirements

Harry T. Oshima, professor of economics, University of Hawaii, Dec. 30, 1974.

Spencer D. Pollard, economics department, University of Southern California.

*Visiting Foreign Scholars**

Geoffrey M. Heal, University of Sussex, England: visiting professor of economics, Yale University, fall 1975.

David F. Hendry, London School of Economics: visiting associate professor of economics, Yale University, fall 1975.

Hiromi Ishigaki: Fulbright-Hays scholar, American Graduate School of International Management, Sept. 1, 1975.

Hartmut Keil, University of Munich: visiting assistant professor, University of Wisconsin-Madison, 1975–76.

Rolf R. Mantel, University of Buenos Aires and Institute Torcuato Di Tella, Argentina: visiting professor of economics, Yale University, July 1975.

Robert Marris, University of Cambridge: visiting professor, department of political economy, Johns Hopkins University, Sept. 1–Dec. 31, 1975.

Ana Maria Martirena-Mantel, University of Buenos Aires and Institute Torcuato Di Tella, Argentina: visiting professor of economics, Yale University, July 1975.

Mihai N. Tarca, University of Iasi, Romania: visiting lecturer, University of Wisconsin-Madison, 1975–76.

Promotions

Bernard E. Anderson: associate professor of industry, Wharton School, University of Pennsylvania, July 1, 1975.

Sven W. Arndt: professor of economics, University of California, Santa Cruz, July 1, 1975.

June Axinn: professor of social welfare, University of Pennsylvania, July 1975.

Sara S. Berry: associate professor, department of economics, Indiana University, July 1, 1975.

Joseph P. Cain: assistant professor of econometrics, department of systems management, Air Force Institute of Technology, School of Engineering, Wright-Patterson Air Force Base, July 15, 1974.

William F. Campbell: professor of economics, Louisiana State University, Aug. 1975.

Robert D. Coston: associate professor of economics, Georgia Southern College, July 1975.

W. Patton Culbertson: associate professor of economics, Louisiana State University, Aug. 1975.

Emit B. Deal: associate professor of economics, Georgia Southern College, July 1975.

Moheb A. Ghali: professor of economics, University of Hawaii, July 1, 1975.

John M. Godfrey: senior financial economist, Federal Reserve Bank of Atlanta, May 1, 1975.

Patrick J. Gormely: associate professor, department of economics, Kansas State University, fall 1975.

Charles P. Gratto: professor of economics, Iowa State University, July 1, 1975.

Kenneth V. Greene: associate professor of economics, State University of New York, Binghamton, Sept. 1, 1975.

David H. Hammond: associate professor of economics and extension resource development specialist, Iowa State University, July 1, 1975.

Robert E. Hicks: professor, department of economics, Florida Technological University.

George E. Hoffer: associate professor of economics, Virginia Commonwealth University, July 1975.

Robert C. Hsu: associate professor of economics, Clark University, Sept. 1, 1975.

David B. Johnson: professor of economics and director of Division of Research, Louisiana State University, Aug. 1975.

Richard H. Keehn: associate professor of economics, University of Wisconsin-Parkside, Sept. 1974.

Thomas E. Kennedy: assistant professor, department of economics, Kansas State University, fall 1975.

Kwan S. Kim: associate professor, University of Notre Dame, 1975–76.

Taeho Kim: associate professor, American Graduate School of International Management, Sept. 1, 1975.

Harvey E. Lapan: associate professor of economics, Iowa State University, Sept. 1, 1975.

William H. Leahy: professor, University of Notre Dame, 1975–76.

Youngil Lim: professor of economics, University of Hawaii, July 1, 1975.

John J. McGowan: vice president, Charles River Associates, Inc., Cambridge, Mass.

Louis J. Maccini: associate professor, department of political economy, Johns Hopkins University, July 1, 1975.

Ronald E. Müller: associate professor, department of economics, American University, Sept. 1, 1975.

Phillip J. Nelson: professor of economics, State University of New York, Binghamton, Sept. 1, 1975.

Alan L. Olmstead: associate professor, department of economics, University of California, Davis, July 1, 1975.

Dennis M. O'Toole: associate professor of economics, Virginia Commonwealth University, July 1975.

Russell G. Pounds: associate professor of economics, Iowa State University, July 1, 1975.

Bertrand M. Renaud: professor of economics, University of Hawaii, July 1, 1975.

J. David Richardson: associate professor of economics, University of Wisconsin-Madison.

James A. Richardson: associate professor of economics, Louisiana State University, Aug. 1975.

Louis A. Rose: associate professor of economics, University of Hawaii, July 1, 1975.

Robert Schafer: associate professor of city planning, department of city and regional planning, Harvard University, July 1, 1975.

Stephen A. Schneider: assistant professor of management and industrial relations, department of management, Wharton School, University of Pennsylvania, July 1, 1975.

Wayne J. Shafer: associate professor, department of economics, Wayne State University, Sept. 1975.

Calvin D. Siebert: professor of economics, University of Iowa, Aug. 1975.

John J. Siegfried: associate professor of economics, Vanderbilt University, Sept. 1975.

V. Kerry Smith: professor of economics, State University of New York, Binghamton, Sept. 1, 1975.

William F. Steel: associate professor of economics, Vanderbilt University, Sept. 1975.

Dale Vorderlandwehr: associate professor, American Graduate School of International Management, Sept. 1, 1975.

Robin J. Walther: assistant professor, economics department, University of Southern California.

Gary M. Walton: professor, department of economics, Indiana University, July 1, 1975.

Jane White: associate professor of business administration, Georgia Southern College, July 1975.

Norman K. Womer: associate professor of economic analysis, department of systems management, Air Force Institute of Technology, School of Engineering, Wright-Patterson Air Force Base, Jan. 1, 1975.

Administrative Appointments

Robert E. Baldwin: chairman, department of economics, University of Wisconsin-Madison.

Whewon Cho: chairman, department of economics and finance, Tennessee Technological University, Sept. 1, 1975.

Clifford D. Clark: president, State University of New York, Binghamton, Mar. 26, 1975.

Robert D. Coston: head, department of economics, Georgia Southern College, Sept. 1975.

John V. Craven: chairman, economics department, Middlebury College, Aug. 1975.

Eugene A. Diulio: chairman, department of economics, Fordham University.

Marshall Geer III: dean of faculty, American Graduate School of International Management, Sept. 1, 1975.

Robert E. Herman: chairman, department of economics, Nassau Community College, Sept. 1975.

John E. La Tourette: acting provost for graduate studies and research, State University of New York, Binghamton, June 2, 1975.

Charles L. Leven: chairman, department of economics, Washington University (St. Louis), July 1, 1975.

Alvin M. Marks: chairman, department of world business, American Graduate School of International Management, Sept. 1, 1975.

John A. Sawyer: director, Institute for the Quantitative Analysis of Social and Economic Policy, University of Toronto, July 1, 1975.

Howard M. Wachtel: chairman, department of economics, American University, June 1, 1975.

William H. Waldorf: acting chairman, department of economics, State University of New York, Binghamton, June 2, 1975.

Ray M. Ware: chairman, department of administration and economics, Transylvania University.

Frederick S. Weaver: dean, School of Social Science, Hampshire College, July 1, 1975.

Murray L. Weidenbaum: director, Center for the Study of American Business, Washington University (St. Louis), Jan. 1, 1975.

Charles K. Wilber, American University: chairman, department of economics, University of Notre Dame, fall 1975.

Appointments

Morris Allen: associate professor of management, American Graduate School of International Management, Sept. 1, 1975.

Robert J. Barro, University of Chicago: professor of economics, University of Rochester, July 1, 1975.

Michael C. Barth, Department of Health, Education, and Welfare: visiting associate professor of economics; research associate, Institute for Research on Poverty, University of Wisconsin-Madison.

Romeo M. Bautista: visiting research fellow, National Bureau of Economic Research, Aug. 1975.

Yoav Benari: economist, Foreign Research Division, Federal Reserve Bank of New York, Sept. 1975.

Lynn Bergold, Southwestern Publishing Co.: special lecturer, North Carolina State University, fall 1975.

H. H. Boggs: assistant professor of world business, American Graduate School of International Management, Sept. 1, 1975.

W. Richard Bossert: assistant professor of world business, American Graduate School of International Management, Sept. 1, 1975.

Ivy Broder: assistant professor, department of economics, American University, Sept. 1, 1975.

Albert E. Burger, Federal Reserve Bank of St. Louis: adjunct professor, St. Louis University, Sept. 1975.

Keith M. Carlson, Federal Reserve Bank of St. Louis: adjunct professor, St. Louis University, Jan. 1976.

Leonard A. Carlson, Occidental College: assistant professor of economics, Emory University, Sept. 1975.

Jens P. Christiansen: lecturer in economics, University of California, Santa Cruz, July 1, 1975.

Robert W. Clarke, University of Houston: associate professor of accounting, Rice University, July 1975.

Kenneth W. Clarkson, University of Virginia: professor of economics, Center for Studies in Law and Economics, University of Miami, and University of Miami Law School, Aug. 1975.

Philip R. P. Coelho, Washington State University: visiting associate professor of economics, University of Hawaii, 1975–76.

Anna M. Craig: visiting assistant professor, Ohio State University.

Sheldon Danriger: lecturer, department of economics; research associate, Institute for Research on Poverty, University of Wisconsin-Madison.

Louis De Alessi, George Washington University: visiting professor of economics, Center for Studies in Law and Economics, University of Miami, Aug. 1975.

Rajat Deb, London School of Economics: visiting assistant professor of economics, Southern Methodist University.

Thomas Dernberg: professor, department of economics, American University, Sept. 1, 1975.

Timothy A. Deyak, State University of New York, Binghamtom: assistant professor of economics, Auburn University, Sept. 1975.

Douglas F. Dowd: visiting professor of economics, University of California, Santa Cruz, Jan. 1, 1976.

James A. Dunlevy, Colby College: assistant professor of economics, Auburn University, Sept. 1974.

James C. Dyer, University of Newcastle, New South Wales: assistant professor of accounting, Rice University, July 1975.

Thomas B. Fomby, University of Missouri, Columbia: assistant professor of economics, Southern Methodist University.

Claude Fongemie, University of Connecticut: instructor, economics department, Middlebury College, Aug. 1975.

Murray Foss: senior research associate, National Bureau of Economic Research, May 1975.

Maxwell J. Fry: associate professor of economics, University of Hawaii, Aug. 1, 1975.

Dilip Kumar Ghosh: visiting assistant professor, Ohio State University.

Reuben Gronau: research associate, National Bureau of Economic Research, Aug. 1975.

Kenneth Guentner, University of Pittsburgh: instructor, economics department, Middlebury College, Aug. 1975.

Thomas D. Hall: assistant professor of economics, University of Hawaii, Aug. 1, 1975.

Eric A. Hanushek, U.S. Air Force Academy: associate professor of economics, Yale University, Jan. 1975.

Robert F. Hebert, Clemson University: associate professor of economics, Auburn University, Sept. 1974.

Betty C. Heian: visiting assistant professor of economics, University of Hawaii, 1975–76.

Robert S. Herren: visiting assistant professor of economics, Vanderbilt University, 1975–76.

James D. Hess: assistant professor, department of economics, University of Southern California.

Martha S. Hill: lecturer, department of economics, Eastern Michigan University, 1975–76.

Randall G. Holcombe: visiting assistant professor of economics, Vanderbilt University, 1975–76.

Karen C. Holden: lecturer, department of economics, University of Wisconsin-Madison.

David L. Horner, Wayne State University: senior research economist, Mathematica, Inc., June 1975.

Ethel B. Jones, University of Georgia: professor of economics, Auburn University, Sept. 1975.

Marcus Jones: economist, Market Statistics Division, Federal Reserve Bank of New York, July 21, 1975.

Barbara H. Kehrer, Center for Health Services Research and Development, A.M.A.: senior research economist, Mathematica, Inc., June 1975.

William A. Kelly, Jr., University of North Carolina, Chapel Hill: assistant professor of economics, Rice University, July 1975.

Stuart H. Kerachsky, University of Wisconsin: researcher, Mathematica, Inc., Dec. 1974.

David T. King: economist, Foreign Research Division, Federal Reserve Bank of New York, July 1, 1975.

Nicholas R. Lardy, University of Michigan: assistant professor of economics and East Asian studies, Yale University, July 1975.

Ross La Roe: instructor, department of economics, American University, Sept. 1, 1975.

William J. Lawrence, New York University: researcher, Mathematica, Inc., July 1974.

Chung H. Lee, Miami University: visiting professor of economics, University of Hawaii, 1975–76.

Wassily Leontief, Harvard University: professor, department of economics, New York University, Sept. 1, 1975.

Ben W. Lewis: visiting professor of economics, Virginia Commonwealth University, spring 1975.

James E. Long, Florida State University: assistant professor of economics, Auburn University, Sept. 1974.

Roland N. McKean, University of Virginia: scholar-in-residence, Center for the Study of American Business; adjunct visiting professor of economics, Washington University (St. Louis), spring semester 1976.

Robert J. McMahon: professor of world business, American Graduate School of International Management, Feb. 1976.

William K. MacReynolds: visiting assistant professor, Ohio State University.

James R. Marchand, California State College, Northridge: assistant professor of economics, Virginia Commonwealth University, Aug. 1975.

Karen Margolis: instructor, North Carolina State University, fall 1975.

Donald L. Martin, University of Virginia: research professor of economics, Center for Studies in Law and Economics, University of Miami, Aug. 1975.

Lawrence C. Marsh: assistant professor of economics, University of Notre Dame, fall 1975.

Andrew W. Mason: assistant professor of economics, University of Hawaii, Aug. 1, 1975.

Rebecca Maynard, University of Wisconsin: researcher, Mathematica, Inc., Dec. 1974.

Leon A. Mears: visiting professor, department of economics, University of Wisconsin-Madison.

Oey Astra Meesook: visiting research fellow, National Bureau of Economic Research, Oct. 1975.

Charles C. Metcalf, University of Wisconsin: principal researcher, Mathematica, Inc., June 1975.

John G. Metcalf: assistant professor of economics, Louisiana State University, Aug. 1975.

Laurence Meyer: economist, Money and Finance Division, Federal Reserve Bank of New York, May 29, 1975.

Marc Miles: economist, Balance of Payments Division, Federal Reserve Bank of New York, Sept. 1, 1975.

James L. Mills: assistant professor of economics, American Graduate School of International Management, Sept. 1, 1975.

John A. Miranowski: assistant professor, department of economics, Iowa State University, Apr. 21, 1975.

Hosseine Morewedge, Long Island University: visiting professor, department of economics, University of Southern California, 1974–75.

Warren E. Moskowitz: economist, Balance of Payments Division, Federal Reserve Bank of New York, May 29, 1975.

Michael M. Murphy, University of Chicago: assistant professor of economics, Southern Methodist University.

William F. Murphy: instructor of business administration, Transylvania University.

David J. Nicol, John Carroll University: assistant professor of finance, Cleveland State University, Sept. 1975.

James O'Connor, Pennsylvania State University: visiting assistant professor, University of Iowa, Aug. 1975.

James C. Ohls, Princeton University: senior research economist, Mathematica, Inc., July 1975.

Carol Oliven, ACT: visiting assistant professor, University of Iowa, Aug. 1975.

Randall J. Olsen, University of Chicago: lecturer in economics, Yale University, July 1975.

Jerome Paige: instructor, department of economics, American University, Sept. 1, 1975.

Donald O. Parsons: visiting research fellow, National Bureau of Economic Research, Sept. 1975.

David W. Patterson: instructor of business administration and hotel and restaurant administration, Transylvania University.

Gerald Peabody: visiting associate professor, department of economics, American University, Sept. 1, 1975.

Andrew Policano: assistant professor of economics, University of Iowa, Aug. 1975.

Walter J. Primeaux, University of Mississippi: professor of business administration, University of Illinois, Urbana, fall 1975.

James Ragan: economist, Business Conditions Division, Federal Reserve Bank of New York, Oct. 1975.

Wallace Reed: assistant professor of accounting, American Graduate School of International Management, Sept. 1, 1975.

Harry W. Richardson: professor, economics and urban and regional planning departments, University of Southern California.

Irving Rottenberg: senior research associate, National Bureau of Economic Research, May 1975.

Daniel L. Rubinfeld: research associate, National Bureau of Economic Research, Sept. 1975.

Javier Ruiz-Castillo: visiting assistant professor, University of Iowa, Aug. 1975.

Richard P. Saba, Texas A&M University: assistant professor of economics, Auburn University, Sept. 1974.

T. Paul Schultz, University of Minnesota: professor of economics of population, Yale University, July 1975.

Steven A. Seelig, Federal Reserve Bank of New York: assistant professor of economics, Fordham University, Sept. 1975.

Jennifer G. Sharpley: lecturer, department of economics, University of Wisconsin-Madison.

Russell Sheldon: economist, Business Conditions Division, Federal Reserve Bank of New York, Sept. 1975.

Parthsarathi Shome: assistant professor, department of economics, American University, Sept. 1, 1975.

Allen Sinai, University of Illinois at Chicago Circle: visiting associate professor, Sloan School, Massachusetts Institute of Technology, Sept. 1975.

Charles Sivesind: economist, Business Conditions Division, Federal Reserve Bank of New York, July 1, 1975.

Robert Sleeper: economist, Foreign Research Division, Federal Reserve Bank of New York, Sept. 1, 1975.

Marcellus S. Snow: assistant professor of economics, University of Hawaii, Sept. 1, 1974.

James B. Stewart: assistant professor of economics and director of black studies, University of Notre Dame, fall 1975.

Stephen H. Strand: visiting assistant professor of economics, Vanderbilt University, 1975–76.

Jack P. Suyderhoud, Advisory Committee on Intergovernmental Relations: assistant professor, department of economics, Eastern Michigan University, 1975–76.

Leigh Tesfatsion: instructor, economics department, University of Southern California.

Francisco Thoumi: research associate, National Bureau of Economic Research, Aug. 1975.

John Vahaly, Jr.: visiting assistant professor of economics, Vanderbilt University, 1975–76.

Kenneth Warner: visiting scholar, National Bureau of Economic Research, July 1975.

Kenneth Wolpin: senior research analyst, National Bureau of Economic Research, Sept. 1975.

Steven A. Wood, Claremont Graduate School: assistant professor, department of economics, Eastern Michigan University, 1975–76.

John Woodbury: economist, Balance of Payments Division, Federal Reserve Bank of New York, Aug. 1975.

Brian D. Wright, Harvard University: lecturer in economics, Yale University, July 1975.

Lorene Y. L. Yap, State University of New York, Binghamton: visiting assistant professor of economics, University of Hawaii, 1975–76.

John M. Yinger: lecturer, department of economics; research associate, Institute for Research on Poverty, University of Wisconsin-Madison.

Soogil Young: visiting instructor, Ohio State University.

David Zimmerman, Temple University: researcher, Mathematica, Inc., July 1975.

Burton Zwick: economist, Money and Finance Division, Federal Reserve Bank of New York, June 1975.

Leaves for Special Appointments

Harold H. Barnett, Washington University (St. Louis): fellow, Woodrow Wilson Center, Smithsonian Institute, Jan. 1, 1976.

Nancy Barrett, American University: deputy assistant director, Congressional Budget Office, June 30, 1975.

Sara S. Berry, Indiana University: visiting associate professor, economics and African history, Boston University, Sept. 1975.

Stanley W. Black, Vanderbilt University: research fellow, Institute for International Economic Studies, University of Stockholm, July 1, 1975–July 30, 1976.

Rudolph C. Blitz, Vanderbilt University: Fulbright fellowship, Institut für Industrie- und Verkehrspolitik, Universität Bonn, Germany, 1975–76.

Nicholas Dewitt, Indiana University: manpower planning adviser, National Manpower Board, Federal Ministry of Economic Development and Reconstruction, government of Nigeria, Aug. 1975.

A. Ronald Gallant, North Carolina State University: visiting assistant professor of business, University of Chicago, 1975–76.

Bruce L. Gardner, North Carolina State University: senior staff economist, President's Council of Economic Advisers, Aug. 1, 1975–July 31, 1976.

R. Jeffery Green, Indiana University: senior staff economist, President's Council of Economic Advisers, Jan. 1975.

Edmond Q. Haggart, Kansas State University: Staff of Joint Congressional Committee on the Budget, June 1975.

George Iden, American University: senior analyst, Congressional Budget Office, May 15, 1975.

James R. Jeffers, University of Iowa: USAID-Korea, May 1975.

Hyman Joseph, University of Iowa: Indiana University, Sept. 1975.

Kwan S. Kim, University of Notre Dame: Rockefeller visiting senior lecturer, University of Dar Es Salaam, 1975–76.

John R. Moroney, Tulane University: visiting professor of economics, Massachusetts Institute of Technology, 1975–76.

Seiji Naya, University of Hawaii: Rockefeller award, visiting professor of economics, Thammasat University, Bangkok, 1975–76.

Edwin G. Olson, Kansas State University: temporary consultant to the Washington State Department of Revenue, Sept. 1975.

Saul Pleeter, Indiana University: economist, U.S. Department of Labor, Washington, D. C., Sept. 1975.

Gordon C. Rausser, Iowa State University: visiting professor, Harvard University, Sept. 1, 1975.

Dominick Salvatore, Fordham University: Fulbright research/lecturer, Italy, 1975–76.

Gian S. Sahota, Vanderbilt University: visiting professor of economics, Fundação Instituto de Pésquisas Economicas (FIPE), University of São Paulo, Brazil, fall semester 1975.

Edward Shapiro, University of Toledo: visiting professor of economics, East Anglia University, England, Sept.–Dec. 1975.

John J. Siegfried, Vanderbilt University: economist, Federal Trade Commission, spring semester 1976.

Gary H. Stern: visiting assistant professor, Washington University (St. Louis), Sept. 1975–June 1976.

Gary M. Walton, Indiana University: visiting professor, department of economics, University of California, Berkeley, Aug. 1975.

James Weaver, American University: faculty, Development Studies Program, Agency for International Development, May 1, 1975.

Kenneth J. White, Rice University: visiting assistant professor, University of British Columbia, 1975–76.

Resignations

Ruben D. Almonacid, Vanderbilt University, May 15, 1975.

Charles W. Bischoff, Yale University: International Business Machines, July 1975.

Thomas F. Dernburg, Oberlin College: American University, Sept. 1, 1975.

Allan C. Deserpa, Louisiana State University, Dec. 1974.

Peter T. Gottschalk, University of California, Santa Cruz, July 1, 1975.

H. Robert Heller, University of Hawaii, June 31, 1975.

Treena K. Levins, Iowa State University, Mar. 14, 1975.

Edwin J. Longfellow, Iowa State University, May 31, 1975.

G. S. Maddala, University of Rochester: University of Florida, Sept. 1, 1975.

Carl W. O'Connor, Iowa State University: Oregon State University, July 18, 1975.

Roy J. Ruffin, University of Iowa: Carlton University, Aug. 1975.

Kenneth R. Smith, University of Wisconsin-Madison: Northwestern University, June 1975.

Frederick T. Sparrow, Johns Hopkins University, Feb. 1, 1975.

Ross M. Starr, Yale University: University of California, Davis, Jan. 1975.

Joseph A. Swanson, University of Iowa: Northwestern University, Aug. 1975.

Darwin Wassink, University of Iowa: Development Assistance Corp., June 1975.

Charles K. Wilber, American University: University of Notre Dame, Sept. 1, 1975.

Michael Zubkoff, Dartmouth Medical School and Dartmouth University, Sept. 1975.

Miscellaneous

Rawle Farley: appointed member, State University of New York University Commission on Purposes and Priorities.

◆

NOTE TO DEPARTMENTAL SECRETARIES AND EXECUTIVE OFFICERS

When sending information to the *Review* for inclusion in the Notes Section, please use the following style:

A. Please use the following categories:

1—Deaths
2—Retirements
3—Foreign Scholars (visiting the USA or Canada)
4—Promotions
5—Administrative Appointments
6—New Appointments
7—Leaves for Special Appointments (NOT Sabbaticals)
8—Resignations
9—Miscellaneous

B. Please give the name of the individual (SMITH, John W.), his present place of employment or enrollment: his new title (if any), his next place of employment (if known or if changed), and the date at which the change will occur.

C. Type each item on a separate 3×5 card, and please do not send public relations releases.

D. The closing dates for each issue are as follows: *March*, November 1; *June*, February 1; *September*, May 1; *December*, August 1.

This announcement supersedes and replaces a letter which was sent annually from the managing editor's office. All items and information should be sent to the Assistant Editor, *American Economic Review*, Box Q, Brown University, Providence, Rhode Island 02912.

SEVENTY-SECOND LIST OF DOCTORAL DISSERTATIONS IN POLITICAL ECONOMY IN AMERICAN UNIVERSITIES AND COLLEGES

The present list specifies doctoral degrees conferred during the academic year terminating June 1975. Abstracts will no longer be printed, as they are published by University Microfilms, Ann Arbor, Michigan.

General Economics; including Economic Theory, History of Thought, Methodology, Economic History, and Economic Systems

BURTON A. ABRAMS, Ph.D. Ohio State 1974. An economic theory of lobbying: A case study of the *U.S.* banking industry.

FRANCIS H. ACKERMAN, Ph.D. Harvard 1975. Riots, populism, and nonindustrial labor: A comparative study of the political economy of the urban crowd.

SYED M. AHSAN, Ph.D. McMaster 1974. Taxation and behavior under uncertainty.

JUDITH A. ALEXANDER, Ph.D. Simon Fraser 1974. Some aspects of consumer related behavior when goods are interrelated.

ROBERT C. ALLEN, Ph.D. Harvard 1975. International competition and the growth of the British iron and steel industry: 1880–1913.

RALPH V. ANDERSON, Ph.D. North Carolina 1974. Labor utilization and productivity, diversification and self-sufficiency, southern plantations, 1800–1840.

ROBERT B. ARCHIBALD, Ph.D. Purdue 1974. Uncertainty, atomistic competition, and excess supply of labor.

WILLIAM M. BAILEY, Ph.D. Maryland 1973. The price revolution, population growth, and agricultural change in Tudor England.

WILLIAM G. BAKER, Ph.D. Nebraska 1975. A business history of Security Mutual Life Insurance Company.

RAYMOND BALLARD, Ph.D. Texas A&M 1974. The demand for nonmarket time.

GERRIT BILKES, Ph.D. State University of New York (Stony Brook) 1974. Macrotheoretic implications of the lot size model for inventories.

CRAIG BOLTON, Ph.D. Texas A&M 1974. The British historical school in political economy: Heterodox meta-economics in the Victorian age.

RHEAD S. BOWMAN, Ph.D. Utah 1975. Jevons on theory and social change.

JAMES W. BOYD, Ph.D. California (Santa Barbara) 1974. Consumer bid transactions with risk.

GERARD F. X. BROWN, Ph.D. Georgetown 1975. Indeterminism and the diffusion index.

JOHN B. BRYANT, Ph.D. Carnegie-Mellon 1975. Inventory behavior.

GERARD BUTTERS, Ph.D. Chicago 1975. Equilibrium price distributions and the economics of information.

JOSEPH P. CAIN, Ph.D. Southern Methodist 1974. An economic analysis of the determination of the quantity and quality of children by the household.

XAVIER CALSAMIGLIA, Ph.D. Minnesota 1975. On the possibility of informational decentralization in nonconvex environments.

JOHN S. COOK, Ph.D. Purdue 1974. Monopolistic solutions in general equilibrium theory.

RICHARD T. CURTIN, Ph.D. Michigan 1975. Perceptions of distributional equity: Their economic bases and consequences.

ANDREW J. DANE, Ph.D. Oklahoma. The industrial sectors of East Germany and West Germany.

JAMES A. DOILNEY, Ph.D. Maryland 1974. Equity, efficiency, and intertemporal resource allocation decisions.

EVAN J. DOUGLAS, Ph.D. Simon Fraser 1974. Edgeworth models of the firm.

JOHN T. DOYLE, Ph.D. Illinois 1975. Cooperation and competition in communal organization.

JAMES T. DYAL, Ph.D. Illinois 1975. Efficiency of investment in graduate education.

JOHN L. EATWELL, Ph.D. Harvard 1975. Scarce and produced commodities: An examination of some fundamentals in the theory of value, with particular reference to the works of Ricardo and Walras.

JOHN C. ECKALBAR, Ph.D. Colorado 1975. Imperfect market information in partial and general disequilibrium.

BENJAMIN EDEN, Ph.D. Chicago 1975. Aspects of uncertainty in simple monetary models.

EVERETT M. EHRLICH, Ph.D. Michigan 1975. The distribution and effects of exogenous disruptions of life cycle plans among labor force participants, 1967–1972.

JOHN B. ELSTROTT, Ph.D. Colorado 1975. Philosophical foundations of liberalism.

RICHARD W. ENGLAND, Ph.D. Michigan 1974. Capital, economic growth, and environmental pollution.

GEORGE M. FEIGER, Ph.D. Harvard 1975. Anticipations and information in speculative markets.

ALEXANDER J. FIELD, Ph.D. California (Berkeley) 1975. Educational reform and manufacturing development in mid-nineteenth century Massachusetts.

CHARLES C. FISCHER, Ph.D. Washington State 1975. Limitations of the equilibrium method in economics.

ROBERT E. FORSYTHE, JR., Ph.D. Carnegie-Mellon 1975. Essays on the theory of incomplete markets.

CHARLES G. GEISS, Ph.D. North Carolina 1975. Consistent price indices.

LEONARD F. GIESECKE, Ph.D. Texas (Austin) 1975.

History of American economic policy in the Philippines during the American colonial period: 1900–1935.

EILEEN C. GRAM, Ph.D. Wisconsin (Madison) 1975. How to raise women's wage—an evaluation of a women's liberation proposal within the context of a general equilibrium model.

SANFORD GROSSMAN, Ph.D. Chicago 1975. Essays on rational expectations, the informational role of futures markets, and equilibrium Bayesian experimentation.

CARL M. GUELZO, Ph.D. Maryland 1974. An institutionalist reconstruction of the theory of the firm.

MILTON Y. HARRIS, Ph.D. Chicago 1974. Optimal planning under transaction costs and general equilibrium.

HIROAKI HAYAKAWA, Ph.D. Michigan 1974. Three essays on consumer choice.

RONALD A. HEINER, Ph.D. California (Los Angeles) 1975. The exchange of collective goods and the existence of generalized competitive equilibrium.

WILLIAM L. HELKIE, Ph.D. Purdue 1974. Studies in applied stochastic microeconomics.

NEAL O. HIGGINS, Ph.D. Nebraska 1974. The early pension plans of the Baltimore and Ohio and the Pennsylvania Railroads, 1880–1937.

JOSEPH P. HUGHES, Ph.D. North Carolina 1974. Two studies of incomplete optimization in economic theory.

TATSURO ICHIISHI, Ph.D. California (Berkeley) 1975. Some generic properties of smooth economies.

FREDERICK S. INABA, Ph.D. California (Berkeley) 1975. Pure exchange equilibrium of a dynamic economy with stochastic endowments.

CHARLES INGENE, Ph.D. Brown 1975. An integration of location and central places theories, given industrial interaction.

RICHARD IPPOLITO, Ph.D. Chicago 1974. The division of labor in the firm.

JAMES S. JORDAN, Ph.D. Northwestern 1975. Temporary competitive equilibrium and the existence of self-fulfilling expectations.

AMOZ KATS, Ph.D. Minnesota 1974. Some models of sequential consumption decisions under uncertainty.

JEAN-JACQUES M. LAFFONT, Ph.D. Harvard 1975. Essays in economics of uncertainty: Information acquisition and instrument dependent randomness.

CHARLES E. LAMBERTON, Ph.D. Iowa State 1975. Localized technical progress in a dynamic theory of the firm.

DENNIS R. LAVIGNE, Ph.D. Illinois 1975. The steady-state growth of the public sector.

JOHN C. LEFGREN, Ph.D. Utah 1975. Banking in the early industrialization of Finland, 1860–1914.

CHARLES LIEBERMAN, Ph.D. Pennsylvania 1974. An investigation of the short-run effects of capital gains on household consumption and saving.

IAN McDONALD, Ph.D. Simon Fraser 1974. A Keynesian theory of the rate of profit.

GORDON C. McMEEKIN, Ph.D. Wayne State 1974. An empirical test of alternative estimation techniques for the consensus and coalition approaches to the committee decision problem.

JAMES G. MACKINNON, Ph.D. Princeton 1975. Economic general equilibrium models: Essays on computation, existence, and estimation.

MARCO MARTINS, Ph.D. Chicago 1975. The private and the social optimum quantity of public debt in a pure consumption-loans model.

GERALD L. MILLER, Ph.D. Notre Dame 1975. Keynesian economics as a "Scientific Revolution."

JOSEPH MINARIK, Ph.D. Yale 1974. A microanalysis of the size distribution of income.

TAPAN MITRA, Ph.D. Rochester 1975. Efficient capital accumulation in infinite horizon models.

JOEL MOKYR, Ph.D. Yale 1974. Industrial growth and stagnation in the low countries, 1795–1850.

SCOTT J. MOSS, Ph.D. New School 1975. On the logic of capitalist accumulation.

EDWARD J. MULHOLLAND, Ph.D. Boston College 1974. An economic analysis of the political behavior of the British Labor Party.

PAUL G. MUNYON, Ph.D. Harvard 1975. A reassessment of New England agriculture in the last thirty years of the nineteenth century: New Hampshire, a case study.

ALLEN R. NEWMAN, Ph.D. North Carolina 1975. An analysis of German internal migration, 1880–1900.

ELI M. NOAM, Ph.D. Harvard 1975. The criminal justice system: An economic analysis of benefits and interrelations.

JAMES F. O'CONNOR, Ph.D. Minnesota 1975. Multiperiod production under uncertainty.

GERALD P. O'DRISCOLL, JR., Ph.D. California (Los Angeles) 1974. Economics as a coordination problem: The contributions of Friedrich A. Hayek.

MARVIN M. OTT, Ph.D. California (Los Angeles) 1975. The consumption-investment decision under uncertainty.

JAMES V. PINTO, Ph.D. Oklahoma. Background and development of location theory: An essay in the history of economic thought.

RICHARD W. T. POMFRET, Ph.D. Simon Fraser 1974. The introduction of the mechanical reaper in Canada 1850–70: A case study in the diffusion of embodied technical change.

ROBERT C. POSATKO, Ph.D. Pennsylvania State 1975. Lifetime parent-to-offspring wealth transfers and utility interdependence.

PATRICIA QUICK, Ph.D. Harvard 1975. Education and industrialization: Elementary education in nineteenth century England and Wales.

RAMA RAMACHANDRAN, Ph.D. Brown 1975. Long-run production function and endogenous technical progress—a neoclassical approach.

WILLIAM S. REECE, Ph.D. Washington (St. Louis) 1975. A disequilibrium model of wage dynamics.

THOMAS P. REINWALD, Ph.D. Illinois 1973. The development and methodology of imperfect and monopolistic competition.

GEORGE RHODES, Ph.D. Texas A&M 1974. An analysis of income, spending, and labor supply in a token economy.

DAVID L. ROBERTS, Ph.D. Duke 1975. Keynes and the system of the Keynesians.

SERGIO ROCA, Ph.D. Rutgers 1975. Economic structure and ideology in socialist Cuba.

THOMAS ROMER, Ph.D. Yale 1974. Social choice criteria and the properties of a simple income tax.

TERRY L. RUSH, Ph.D. Oklahoma. The Keynesian-monetarist debate: An analysis and synthesis.

TERRANCE C. I. RYAN, Ph.D. Massachusetts Institute of Technology 1975. The economics of trading in slaves.

RICHARD SABA, Ph.D. Texas A&M 1974. Information cost and the theory of discrimination.

ANNA M. SACHKO, Ph.D. Columbia 1975. An analysis of the demand for life insurance.

ERNESTO F. SANZ, Ph.D. Boston College 1975. Polity primacy and microeconomic structures in the economic development of Taisho Japan.

MARK D. SCHMITZ, Ph.D. North Carolina 1974. Economic analysis of antebellum sugar plantations in Louisiana.

JOHN J. SEATER, Ph.D. Brown 1975. A unified model of consumption, labor supply, and job search.

GORDON H. SELLON, JR., Ph.D. Michigan 1974. A want-hierarchy approach to the theory of consumer behavior.

MARVIN SHAFFER, Ph.D. British Columbia 1974. The role of competition in macro models.

IRWIN A. SHAPIRO, Ph.D. Clark 1974. A test of the effect of windfall receipts on consumption.

STEVEN M. SLUTSKY, Ph.D. Yale 1975. Majority voting and the allocation of public goods.

JON C. SONSTELIE, Ph.D. Northwestern 1975. Public goods and property values.

DAVID H. SWINTON, Ph.D. Harvard 1975. The logic of economic discrimination in noncompetitive labor markets.

MO-YIN TAM, Ph.D. State University of New York (Stony Brook) 1974. On the role of "information flows" in the theory of optimal policies.

ARVIND VIRMANI, Ph.D. Harvard 1975. A dynamic theory of the firm: Expectations, investment, and growth.

MICHAEL L. VISSCHER, Ph.D. Virginia 1975. The value of consumers' waiting time in the supply of goods.

MARK S. WALDMAN, Ph.D. Florida 1974. Galbraith's paradigm: A case study in scientific revolution.

PETER G. WARR, Ph.D. Stanford 1974. The theory of shadow pricing.

FRANK WESTHOFF, Ph.D. Yale 1974. The theory of local public goods.

HUGH R. WILLS, Ph.D. California (Berkeley) 1975. Vintage capital and the theory of the firm.

RICHARD J. WILTGEN, Ph.D. Illinois 1974. Marxism and population: A reappraisal.

DANIEL WISECARVER, Ph.D. Chicago 1974. Three essays in the theory of applied welfare economics.

GLENN A. WITHERS, Ph.D. Harvard 1975. The political economy of the volunteer military.

GARY W. YOHE, Ph.D. Yale 1975. A comparison of price controls and quantity controls under uncertainty.

JEFFREY T. YOUNG, Ph.D. Colorado 1975. The rehabilitation of the classical theory of value in historical perspective.

Economic Growth and Development; including Economic Planning Theory and Policy, Economic Fluctuations and Forecasting

DOUGLAS ADKINS, Ph.D. California (Berkeley) 1975. Economic planning and the specialized degree output of higher education in the United States.

ABUBAKR BIN AHMAD, Ph.D. Indiana 1974. Development strategies for Ghana: A determination of alternative policies for 1980.

FAKHRUDDIN AHMED, Ph.D. Princeton 1975. Migration and employment in a multisector model: An application to Bangladesh.

LOUKAS I. ANANIKAS, Ph.D. Michigan State 1975. Potential livestock production adjustments on family farms in central Macedonia, Greece.

YIADOM ATTAKONDADU, Ph.D. Michigan State 1974. Economic optima in resource allocation for smallholder subsistence farming in Ghana.

CARLOS A. BAANANTE, Ph.D. North Carolina State 1974. Andean group economic integration: The case of the nitrogenous fertilizer industry.

LLOYD BARON, Ph.D. McGill 1975. The water supply constraint: An evaluation of irrigation projects and their role in the development of Afghanistan.

CIHAN M. BELEN, Ph.D. Connecticut 1975. The role of the state economic enterprises in the development of Turkey.

JEFFREY I. BERNSTEIN, Ph.D. Western Ontario 1975. Essays on the dynamic theory of optimal policy for an open economy.

ASHOK BHARGAVA, Ph.D. Wisconsin (Madison) 1975. The effect of a rise in the rate of interest on corporate behavior: A case study of India.

STEVEN Y. F. BONAWU, Ph.D. Nebraska 1974. The financial sector and economic development: The Liberian case.

FRANCOIS BOURGUIGNON, Ph.D. Western Ontario 1975. Two essays in economic growth theory.

IVY BRODER, Ph.D. State University of New York (Stony Brook) 1974. Statistical analyses of cyclical indicators.

EMIN CARIKCI, Ph.D. New School 1975. Workers' migration to the European Economic Community (EEC) countries and Turkish economic development.

Daniel S. Christiansen, Ph.D. Stanford 1975. Some aspects of the theory of short-run equilibrium.

W. Edmund Clark, Ph.D. Harvard 1974. Socialist development and public investment in Tanzania: 1964–1973.

Harry M. Cleaver, Jr., Ph.D. Stanford 1975. The origins of the Green Revolution.

Arthur L. Coffing, Ph.D. Iowa State 1974. PL-480 imports, the disincentive effect, and implications for development in Turkey.

William E. Conrad, Jr., Ph.D. Massachusetts Institute of Technology 1975. Uncertainty and optimal stabilization policy.

Roy E. J. Culpepper, Ph.D. Toronto 1975. Import substitution, self-reliance, and the choice of techniques: A study of the Indian fertilizer industry.

Gunars Dambe, Ph.D. Iowa State 1974. Planning for more labor intensive and more productive agriculture in Columbia.

Asim K. Dasgupta, Ph.D. Massachusetts Institute of Technology 1975. Two essays on income distribution in a developing economy.

Nicholas J. DeGrazia, Ph.D. Michigan State 1974. Stock markets and development: A case study of Panama.

Lawrence E. Demilner, Ph.D. Southern California 1975. The formation of price expectations and their behavior in wage change and interest rate equations.

Evelyn Hazel Denton, Ph.D. Harvard 1975. Economic determinants of fertility in Jamaica.

Bhupendra M. Desai, Ph.D. Cornell 1974. Relationship of consumption and production in changing agriculture: A study in Surat District, India.

Carl A. Futia, Ph.D. California (Berkeley) 1975. Information costs and boundedly rational behavior: A comparative equilibrium.

Kenneth D. Garbade, Ph.D. Princeton 1975. Discretionary control of aggregate economic activity.

Arthur Gibb, Jr., Ph.D. Michigan 1974. Agricultural modernization, nonfarm employment, and low-level urbanization.

Nicholas P. Glytsos, Ph.D. Syracuse 1974. Dynamic price changes, wages, and unemployment in the urban sector of the less developed countries: An intertemporal cross-sectional econometric model.

Robert E. Gray, Ph.D. Michigan State 1975. Agricultural export potentials and balance-of-payments aspects of the Nigerian economy.

Jean Guertin, D.B.A. Harvard 1974. An evaluation of forecasts by financial analysts.

Lawrence H. Hadley, Ph.D. Connecticut 1975. An international comparative analysis of educational strategies in under-developed countries incorporating rates of return to education.

Khalil Ahmed Hamdani, Ph.D. Georgetown 1975. Risk, social structure, and economic behavior in the peasant economy.

Barry M. Herman, Ph.D. Michigan 1974. Strategies of foreign manufacturing investment in poor countries: A general theory and an application to oil refining in East Africa.

Shigemochi Hirashima, Ph.D. Cornell 1974. Interaction between institutions and technology in developing agriculture: A case study of the disparity problems in Pakistan agriculture.

Joan Hoffman, Ph.D. New School 1974. The interaction of racial discrimination and economic development: A case study of South Carolina: 1960–70.

Sebastiao Jorge Jatoba Bezerra dos Santos, Ph.D. Vanderbilt 1975. Factor price policies, technological change, and employment: The case of the Brazilian northeast manufacturing industry, 1949–1969.

Tim D. Johnston, Ph.D. Iowa State 1974. Income potential of small farms in Guatemala.

Christos Kamenidis, Ph.D. Michigan State 1974. Efficient organization of the livestock meat marketing system in eastern Macedonia, Greece.

Vassilios Kanellakis, Ph.D. Kansas State 1975. International tourism: Its significance and potential as an instrument for the economic development of Greece.

Basant K. Kapur, Ph.D. Stanford 1975. Monetary growth models of less developed economies.

Sahika Kazancigil Kayir, Ph.D. Iowa State 1974. Planning for growth: The case of Turkey.

Byong Yong Kim, Ph.D. Claremont 1975. Agricultural vs. industrial expansion in the development of Korea: A proposal for a program of meat production under government control.

Joong-Woong Kim, Ph.D. Clark 1975. The role of money in economic growth: Some empirical evidence from less developed countries.

Sven B. Kjellstrom, Ph.D. Michigan 1974. The impact of tourism on economic development in Morocco.

Lily Kosiyanon, Ph.D. Oregon 1974. The behavior of household savings in Thailand.

Yuji Kubo, Ph.D. Purdue 1974. Two-sector models of urban concentration and economic development.

Ramesh C. Kumar, Ph.D. Toronto 1975. Development of an open dual economy: A growth-theoretic study.

Yoshimi Kuroda, Ph.D. Stanford 1975. A study of Japanese farm household behavior in production and consumption in the postwar years.

Daniel Landau, Ph.D. Chicago 1974. The business cycle and productivity change.

Charles T. Lawton, Ph.D. Fletcher School 1975. The empirical basis for the concept of over-urbanization: An examination of the demographic and economic patterns for sub-Saharan Africa.

Jeung H. Lee, Ph.D. Michigan State 1975. Projections of product supply and factor demand under structural change for Korean agriculture: A systems simulation approach.

Chandi Charan Maji, Ph.D. Iowa State 1975. Intertemporal allocation of irrigation water in the Mayurakshi Project (India): An application of deterministic and chance-constrained linear programming.

Humphrey H. Milner, Ph.D. Chicago 1975. Puerto Rican growth: A two-stage *CES* production function.

Gouda A. E. Mohamed, Ph.D. McMaster 1974. Sectoral interdependence and Egypt's investment strategy.

Dale L. Moody, Ph.D. Florida 1975. The manufacturing sector in the Republic of Vietnam: Its structure, productivity, and development.

Thomas K. Morrison, Ph.D. Maryland 1974. The performance of less developed countries in exporting manufactures: A cross-country analysis.

Kadima G. Muamba, Ph.D. Florida State 1975. Required tax policy and economic development of Zaire.

George A. Munguia, Ph.D. Fordham 1975. A transitional stage in the agricultural sector: The case of the Dominican Republic.

N. V. R. Murthy, Ph.D. State University of New York (Binghamton) 1975. Potential output and sources of economic growth in the Indian economy: 1950–51 to 1964–65.

Mustapha Kamel Nabli, Ph.D. California (Los Angeles) 1974. An econometric model for Tunisia: The use of a growth and development model for policy simulation and evaluation.

Werner Neu, Ph.D. California (Berkeley) 1975. Model for computer simulation of fiscal, monetary, and balance-of-payments policies in a developing economy: Thailand.

Lu Siong Ng, Ph.D. Iowa State 1975. An income distribution and employment consistency model of the Philippines.

Mieko Nishimizu, Ph.D. Johns Hopkins 1975. Total factor productivity analysis: A disaggregated study of the postwar Japanese economy with explicit consideration of intermediate input.

Diann H. Painter, Ph.D. Indiana 1974. Planning for industrial growth in a colonial and postcolonial setting: The case of Senegal, 1949–1970.

Jong-Heum Park, Ph.D. Oklahoma State 1974. Causes and effects of export instability in a developing country: The case of Pakistan.

Terrance C. Parks, Ph.D. Illinois 1974. The impact of the petroleum industry on the economic development of Libya.

Armando L. Pinell-Siles, Ph.D. Harvard 1975. Economic growth and income distribution in the dual economy.

Mary Pitts, Ph.D. Syracuse 1975. Economic development of Chile, 1925–1968.

R. David Ranson, Ph.D. Chicago 1974. Money, capital, and the stochastic nature of business fluctuations.

P. H. Rayappa, Ph.D. Florida State 1974. Determinants of labor force participation in a developing economy: An analysis for India.

Osmundo E. Reboucas, Ph.D. Harvard 1975. The interregional effects of economic policies: Multisectoral general equilibrium estimates for Brazil.

Alan R. Richards, Ph.D. Wisconsin (Madison) 1975. Accumulation, distribution, and technical change in Egyptian agriculture, 1800–1940.

Neil B. Ridler, Ph.D. Simon Fraser 1974. The relationship between land tenure systems and development goals: A study of the Colombian palm oil industry.

Alfredo Roa, Ph.D. Michigan State 1974. Rural-urban migration and its relation to unemployment in the urban area of Valle del Cauca, Colombia.

Charles R. Roll, Jr., Ph.D. Harvard 1975. The distribution of rural incomes in China: A comparison of the 1930's and 1950's.

Francis W. Rutaiwa, Ph.D. Fordham 1975. Application of incentives for rural development in developing countries: A case study of Tanzania.

Emilio Sacerdoti, Ph.D. Yale 1975. Inflation and rates of return: An equilibrium approach.

Abul H. Sahadat Ullah, Ph.D. Harvard 1974. Farm productivity, returns to scale, and technological change in a traditional agriculture (Bangladesh).

Nicholas P. Sargen, Ph.D. Stanford 1975. "Tractorization" in the United States and its relevance for developing countries.

Bansi Sawhney, Ph.D. George Washington 1975. An analysis of the interrelationship of India's export trade and economic development.

Enrique O. Scala, Ph.D. Minnesota 1974. The efficiency of import substitution in the Argentine automotive industry.

John E. Scarbrough, Ph.D. Minnesota 1974. Some economic effects of customs unions in South America.

Raul Schkolnick, Ph.D. Harvard 1975. The economics of telephone networks: A case study of a developing country.

Soelistyo, Ph.D. Colorado 1975. Creating employment opportunities in the rural areas of East Java.

Mary S. Staig, Ph.D. Texas (Austin) 1975. The historical evolution of rights in land in Bihar, India.

Paul K. Steidlmayer, Ph.D. Stanford 1975. The TaChai model in Chinese agriculture: A study of policy, its implementation and performance (1960–74).

John R. Stewart, Jr., Ph.D. North Carolina 1974. An analysis of the two-gap theory of economic development, with empirical application for the Mexican economy.

Peter Sturm, Ph.D. Yale 1974. A comparison between aggregate production relationships in East and West Germany.

Alexander A. Sunday, Ph.D. Illinois 1975. Estimation of the effectiveness of international tourism promotion by selected countries in American media.

Bunloe Sutharomn, Ph.D. Michigan State 1974. The effect of changing areas and levels of technology of corn and cotton production on the net foreign exchange in Thailand.

CHUMPHOT SUVAPHORN, Ph.D. North Carolina 1975. Model of agricultural development in the less developed country: A study of rice production in Thailand.

GORDON P. TEMPLE, Ph.D. Wisconsin (Madison) 1974. Migration to Jakarta: An empirical search for a theory.

CLEMENS TIBAZAWA, Ph.D. Claremont 1975. Economic impact on Tanzania of the East African common market.

BARBARA TUCKMAN, Ph.D. Florida State 1974. The Green Revolution, agricultural productivity, and the income distribution in Mexico.

RICHARD O. WADA, Ph.D. Hawaii 1975. Impact of economic growth on the size distribution of income: The postwar experience of Japan.

CAM H. WICKHAM, Ph.D. Oregon 1975. Venezuela's royal tobacco monopoly 1779–1810: An economic analysis.

JOHN H. WILHELM, Ph.D. Michigan 1974. The Soviet attempt at regional economic planning: The Sovnarkhoz 1957–1965.

SAMUEL A. WOODS, Ph.D., Harvard 1975. Food cycles and crises: A spectral analytic inquiry into the nature of the world food problem.

JONGTACK YOO, Ph.D. Michigan State 1975. A short- and long-run analysis of the Korean rural demand for food and its implications to agricultural policies.

AMIN ZEWAIL, Ph.D. Fordham 1975. Egypt's foreign borrowing requirements: An import-export approach.

ANDREW S. ZIMBALIST, Ph.D. Harvard 1974. Worker participation in the management of specialized industry: An empirical study of the Chilean experience.

HERBERT F. ZOLLNER, Ph.D. Colorado 1975. Planning for socioeconomic development: Some contributions to increasing its efficiency.

Economic Statistics; including Econometric Methods, Economic and Social Accounting

JAMES R. ALDRICH, Ph.D. Purdue 1974. Foundations for a decision-theoretic approach to scientific inquiry.

ANUPAM BASU, Ph.D. Stanford 1975. Control level determination in regression models.

GARY E. CHAMBERLAIN, Ph.D. Harvard 1975. Unobservables in econometric models.

PETER COUGHLIN, Ph.D. New Mexico 1975. Predictions of input-output analysis.

DENNIS EPPLE, Ph.D. Princeton 1975. An econometric analysis of the supply of petroleum discoveries in the United States.

ROQUE FERNANDEZ, Ph.D. Chicago 1975. Short-run dynamics of output and prices.

SHELBY D. GERKING, Ph.D. Indiana 1975. Some statistical problems in estimating stochastic input-output models.

ROB GERRISTEN, Ph.D. Carnegie-Mellon 1975. Understanding data structure.

ROBERT C. GOLDSTEIN, D.B.A. Harvard 1974. The cost of privacy.

GARY L. GRIEPENTROG, Ph.D. Purdue 1974. The relationship among model misspecification, estimation technique and forecasting in some econometric marketing models: A computer experiment.

ELIZABETH F. GUSTAFSON, Ph.D. North Carolina 1974. Dynamic econometrics models: Dynamic simulation forecasts and testing for stability.

JAMES D. HESS, Ph.D. Massachusetts Institute of Technology 1975. Team decision theory.

STEPHEN J. HUXLEY, Ph.D. California (San Diego) 1975. An interregional input-output model for the state of Queensland, Australia, 1970–71.

RANGANATH K. R. IYENGAR, Ph.D. Wisconsin (Madison) 1974. Toward a complete saving expenditure system with an application to Taiwan data.

RAJ K. JAIN, Ph.D. Purdue 1975. Estimation of a distributed lag model.

WILLIAM J. JOHNSON, JR., Ph.D. Connecticut 1975. A linear expenditure system with random coefficients.

KARIMA A. M. KORAYEM, Ph.D. McMaster 1974. The structure and the development of the financial flows in Egypt during 1952–70.

KAZUTAKA KUNIMOTO, Ph.D. Yale 1975. The contingency-table analysis of international trade flows: An approach in methodology.

FRANCIS R. LAD, Ph.D. Michigan 1974. On the adequacy of natural conjugate distributions for representing prior information.

KAJAL LAHIRI, Ph.D. Rochester 1975. Estimation of econometric models with unobservable variables.

ARTHUR J. MALLOY, Ph.D. State University of New York (Stony Brook) 1974. The production function and heterogeneous labor.

JACK MEYER, Ph.D. Stanford 1975. Stochastic dominance, increasing risk, and risk aversion.

DANIEL F. MOHAN, Ph.D. Rutgers 1975. Alternative specifications of qualitative dependent variable models: A Monte Carlo approach.

FORREST D. NELSON, Ph.D. Rochester 1975. Estimation of economic relationships with censored, truncated, and limited dependent variables.

CHRISTOPHER PIPER, Ph.D. Carnegie-Mellon 1975. Computational studies in optimizing and postoptimizing linear programs in zero-one variables.

HOSSEIN RAZAVI-GHAHFAROKHI, Ph.D. Maryland 1974. Optimal stochastic control of econometric models.

THOMAS C. REIMBOLD, Ph.D. Maryland 1974. Simulation with a dynamic interindustry forecasting model.

AJAY K. SANGHI, Ph.D. State University of New York (Stony Brook) 1974. The determinants of automobile ownership.

FREDERICK W. SIEGEL, Ph.D. Illinois 1974. A mean variance skewness model of security demand.

JAMES STARR, Ph.D. Carnegie-Mellon 1975. Zero-one programming: A partial ordering and its use in the optimization of numerical and non-numerical functions.

Bert M. Steece, Ph.D. Southern California 1974. A decision-theoretic approach to response surface methodology.

Krishnappa Swaminatha, Ph.D. Boston 1975. Causal factors in price movements in India, 1950–51 to 1970–71: An empirical analysis.

William E. Taylor, Ph.D. California (Berkeley) 1975. Linear estimation when the disturbance covariance matrix is unknown.

Noel D. Uri, Ph.D. Illinois 1974. Towards an efficient allocation of electrical energy.

Jon D. Vasche, Ph.D. California (Berkeley) 1975. An econometric analysis of the market demand for mobile home housing.

Stephen Whitelaw, D.B.A. Harvard 1975. The use of mathematical models in organizational problem solving.

Benjamin Wurzburger, Ph.D. Massachusetts Institute of Technology 1974. Long-run properties of the MIT-PENN-SSRC econometric model.

Michael D. Yokell, Ph.D. Colorado 1975. Distribution shares, economic class, and income inequality in the United States: 1922–1971.

Carol Zarley, Ph.D. Kansas 1975. Statistical estimations in a random coefficient errors in variables model.

Monetary and Fiscal Theory, Policy, and Institutions

Mabid Mahmoud Al-Jarhi, Ph.D. Southern California 1975. The optimal supply of money and optimal monetary policies.

John A. Andrulis, Ph.D. Purdue 1975. An optimal tax structure for urban growth.

William T. Baldwin, Ph.D. Kentucky 1975. The desirability and feasibility of initiating explicit interest payments on commercial bank demand deposits.

Anthony J. Barkume, Ph.D. California (Santa Barbara) 1974. Empirical studies on voting behavior on fiscal referenda.

James A. Berkowicz, Ph.D. Maryland 1974. Bank holding company conduct, structural change, and the performance of banks.

Zvi Bodie, Ph.D. Massachusetts Institute of Technology 1975. Hedging against inflation.

Jerry J. Bodily, Ph.D. Purdue 1974. An economic analysis of retaliation and discrimination in the state taxation of life insurance companies.

Walter G. Bolter, Ph.D. Maryland 1973. The theory, development, and application of national economic budgeting with special reference to England, Sweden, and the United States.

Patricia D. Brenner, Ph.D. Stanford 1975. The impact of Federal Reserve regulations on the Eurodollar market.

David W. Butterfield, Ph.D. California (Berkeley) 1975. Money as a productive factor: A study of balances.

Charlie Carter, Ph.D. Illinois 1975. An analysis of intergovernmental expenditure dispersions.

John H. Ciccolo, Jr., Ph.D. Yale 1975. Four essays on monetary policy.

Gary E. Clayton, Ph.D. Utah 1974. The money-as-net-wealth hypothesis.

Thomas E. Cody, Ph.D. Harvard 1974. Evaluation of reserves regimes.

Rex Cottle, Ph.D. Texas A&M 1974. An economic analysis of the federal funds market.

John Z. Drabicki, Ph.D. Purdue 1974. Studies in money, the national debt, and economic growth.

Felix R. Dy Reyes, Ph.D. Iowa State 1974. A test of the direction of causation between money and income in Canada, Japan, and the United States.

James Gale, Ph.D. Iowa 1975. Subsidized capital inflows and regional income distribution.

John M. Gemello, Ph.D. Stanford 1975. Redistribution effects of intergovernmental grants: A study of the educational finance system in California.

Antonio Goncalves, Ph.D. Chicago 1974. The problem of stopping inflation.

Robert A. Gough, Jr., Ph.D. Duke 1975. Intergovernmental grants-in-aid: A general model for assessing the local fiscal effects of the Massachusetts variable-matching public school aid program.

James S. Grichar, Ph.D. Wisconsin (Milwaukee) 1975. The effect of the market value of corporate stocks and bonds on aggregate consumption.

Warner N. Grubb, Ph.D. Harvard 1975. Intergovernmental aid and resource disparities: School finance in Massachusetts.

Charles E. Guy, Ph.D. Florida 1974. Liquidity preference theory: Its relation to manufacturing corporations.

Jack I. Habib, Ph.D. Harvard 1974. Comparing alternative tax-transfer structures.

Theodore Haniotis, Ph.D. Oregon 1975. An aggregate model of monetary behavior in Greece: A frame for a non-linear ratio form money supply hypothesis.

Nozar Hashemazadeh, Ph.D. Virginia Polytechnic Institute 1975. Banks and inflation.

Sydney S. Hicks, Ph.D. Washington (St. Louis) 1975. Interest rates and the controversy over money.

Thomas Hoenig, Ph.D. Iowa State 1974. Commercial banking: Competition and the personal loan market.

Charles F. Hoffman, Ph.D. California (Davis) 1974. A model of the consumer credit industry and consumer durable expenditures.

Stuart G. Hoffman, Ph.D. Cincinnati 1975. Determination of the money multiplier by component ratio estimation.

Robert B. Hool, Ph.D. California (Berkeley) 1975. Money, financial assets, and general equilibrium in a sequential market economy.

Carl M. Hubbard, Ph.D. Texas Tech 1975. Federal Reserve membership, bank size, and bank location: Their effects on selected operating ratios and profitability of Texas commercial banks.

M. Edwin Ireland, Ph.D. Texas Tech 1974. A statistical evaluation of the cause-effect relationship between money and income.

JOHN JAMES, Ph.D. Massachusetts Institute of Technology 1974. The evolution of the national money market, 1888–1911.

DONG-KUEN JEONG, Ph.D. Wayne State 1974. A fiscal policy decision model for the *U.S.* economy: A revealed preference inquiry into the optimal policy solution during the first year of the Eisenhower Administration.

MONZER KAHF, Ph.D. Utah 1975. On recent developments in monetary growth theory.

SUN MYOUNG KAHNG, Ph.D. Indiana 1975. A study of the Korean inflation, 1936–71: A quantitative analysis of the factors determining real output, money *GNP*, and price level.

SUHEIL KAWAR, Ph.D. California (Los Angeles) 1974. The evolution of monetary institutions and the money stock in Lebanon, 1918–68.

SOO-KEUM KIM, Ph.D. Vanderbilt 1975. A study of public expenditure growth among the American states, 1902–71.

LEVIS KOCHIN, Ph.D. Chicago 1975. Judging stabilization policies.

GARY E. KUNDEY, Ph.D. Florida 1975. Determinants of the outstanding volume of dollar bankers' acceptances.

IRWIN D. LANE, Ph.D. California (Berkeley) 1975. Commercial bank loan commitments.

EDWARD F. MCKELVEY, Ph.D. Yale 1975. Bank portfolio decisions and deposit uncertainty: A comparison of alternative rules of allocation.

RICHARD C. MCKIBBIN, Ph.D. California (Berkeley) 1975. The distribution of federal credit subsidy benefits.

PAUL D. MCNELIS, Ph.D. Johns Hopkins 1975. A simulation analysis of monetary policy and macroeconomic adjustment.

ALLAN S. MANDEL, Ph.D. Michigan 1974. Resource allocation inside school districts.

BRIAN A. MARIS, Ph.D. Kansas State 1974. Application of the "St. Louis" short-term macroeconomic forecasting model to Canada.

ANN MARKUSEN, Ph.D. Michigan State 1974. The economics of social class and metropolitan local government.

DAVID K. MEAD, Ph.D. Clark 1974. A model for predicting voters' behavior: A case study of the Massachusetts graduated income tax referendum.

JAMES L. MEDOFF, Ph.D. Harvard 1975. The general equilibrium effects of governmental expenditures on the family distribution of income.

JAMES S. MOORE, Ph.D. Purdue 1974. The asset and liability behavior of large commercial banks: A disaggregated econometric approach.

JOSEPH L. MOORE, Ph.D. Arkansas 1975. A study of relationships between prices and interest rates for the period 1890 through 1970, using selected time intervals.

THOMAS V. MUNZENRIEDER, Ph.D. Minnesota 1974. Fluctuating interest rates and savings and loan profitability.

EDWIN A. MURRAY, D.B.A. Harvard 1974. The implementation of social policies in commercial banks.

CEDRIC D. NATHAN, Ph.D. Simon Fraser 1974. The money supply and interest rates in South Africa 1950–71: An econometric enquiry.

WELDON C. NEILL, Ph.D. Texas (Austin) 1974. Fluctuations in the flow of funds to nonfarm residential mortgage markets as a consequence of monetary policy.

WILLIAM B. NELSON, Ph.D. Rice 1975. Target rate of return pricing and the forward shifting of the corporation income tax in *U.S.* manufacturing.

PAULO NEUHAUS, Ph.D. Chicago 1974. A monetary history of Brazil, 1900–45.

NICHOLAS N. NOE, Ph.D. Indiana 1974. An econometric study of state and local government finance in the state of Indiana.

YOICHI OKITA, Ph.D. Harvard 1974. Domestic and international impacts of Japanese fiscal policy: 1951–1969.

FRANK S. ORLANDO, Ph.D. Purdue 1974. The interest rate behavior of bank reserves and flow-of-funds financial models.

RICHARD P. O'TOOLE, Ph.D. California (San Diego) 1974. Long-run monetary and fiscal policy in a mixed economy.

DIMITRIOS PANAYOTOPOULOS, Ph.D. Virginia Polytechnic Institute 1975. The impact of foreign influences upon *U.S.* money stock components.

PETROS P. PAPAGEORGIOU, Ph.D. Boston 1975. The incidence of taxation and the availability of free higher education in Greece.

JAE-YOON PARK, Ph.D. Indiana 1975. A structural estimation of money demand and supply functions for Korea.

ROBERT K. PERDUE, Ph.D. West Virginia 1974. Evidence on the shifting of property taxes in a regulated industry.

JOSEPH M. PRINZINGER, Ph.D. Georgia State 1974. The effect of membership status in the Federal Reserve System on bank profit: A national study.

RONALD A. RATTI, Ph.D. Southern Methodist 1975. Model of a commercial bank under uncertainty.

DAVID M. REAUME, Ph.D. Harvard 1974. Local income taxes: A theoretical and empirical analysis.

TROND R. REINERTSEN, Ph.D. Utah 1975. Money and financial structure.

SOMCHAI RICHUPAN, Ph.D. Florida 1974. The state and local government budget decision: An application of the linear expenditure model.

TERRY R. RIDGWAY, Ph.D. Arizona 1974. The responsiveness of Nevada state tax revenues to changing economic conditions.

Henri-Paul Rousseau, Ph.D. Western Ontario 1974. Application of the theory of markets to provincial finance: An econometric and empirical analysis of Alberta budgetary realization process.

Richard C. Schiming, Ph.D. Ohio State 1974. The United States on a silver standard, 1873–1900: A counterfactual history.

Maria Schmundt, Ph.D. Yale 1974. National and international tax incidence in a simple income tax.

Howard C. Schneider, Ph.D. Virginia 1974. An economic analysis of selected grant-in-aid programs.

Mary E. Scovill, Ph.D. Illinois 1975. Local taxation in Uganda, Kenya, and Zambia.

Parthasarathi Shome, Ph.D. Southern Methodist 1975. Specific factors and the general equilibrium theory of tax incidence.

Charles M. Sivesind, Ph.D. Iowa State 1975. A dynamic programming simulation of optimal monetary policies designed to stabilize prices and employment.

Charles Smaistrla, Ph.D. Texas A&M 1974. Eurodollars in commercial bank liquidity management.

Dave Spencer, Ph.D. Texas A&M 1974. Transmission of monetary policy.

Larry E. Steinhauer, Ph.D. Chicago 1974. The demand for cash balances by business firms: An empirical investigation.

C. Eugene Steuerle, Ph.D. Wisconsin (Madison) 1975. Expected rates of return on savings portfolios: The variance across socioeconomic classes.

Marie E. Sushka, Ph.D. Georgetown 1975. An econometric model of the money market in the United States, 1823–59.

John E. Sutherland, Ph.D. Virginia Polytechnic Institute 1975. An analysis of the effects of monetary policy on aggregate and sectoral output and prices.

Thomas J. Sweeney, Ph.D. Iowa State 1974. An application of the polynominal distributed lag to monthly, quarterly, and annual data in the demand function for money.

Ronald J. Talley, Ph.D. Johns Hopkins 1975. The relative importance of speculative and precautionary motives for holding money.

Laura D. Tyson, Ph.D. Massachusetts Institute of Technology 1974. Inflation in Yugoslavia, 1962–72: An empirical analysis.

Paul N. Van De Water, Ph.D. Massachusetts Institute of Technology 1975. Property tax incidence in a macroeconometric model.

Jennie G. Versteeg, Ph.D. Clark 1974. A median voter approach to the explanation of variations in local government expenditures.

Kevin E. Villani, Ph.D. Purdue 1974. The role of nonbank financial intermediaries in the capital markets.

Susan L. Wachter, Ph.D. Boston College 1974. Latin American structuralist and monetarist inflation theories: An application to Chile.

Jerome A. Wade, Ph.D. Syracuse 1974. The role of tax base composition in school finance decisions and reform.

Terrence W. West, Ph.D. Connecticut 1975. Fiscal federalism and general revenue sharing: A study of alternative proposals in the Ninety-Second Congress.

Stanley L. Winer, Ph.D. Johns Hopkins 1975. Monetary-fiscal influences in federal states: With application to the postwar Canadian economy.

Allen M. Wyse, Ph.D. Illinois 1974. The incidence of highway taxation.

James A. Xander, Ph.D. Georgia 1975. Vault cash and the role of commercial banks and regional economic development—United States 1870–1913.

Peter S. Yun, Ph.D. Georgia 1975. An investigation of the attitudes of the Federal Reserve open market committee to wages and employment, 1951–67.

Howard R. Zacks, Ph.D. Wayne State 1975. Fiscal policy and aggregate demand in a monetary economy.

International Economics

Robert A. Baada, Ph.D. Wisconsin (Madison) 1974. *U.S.* commercial policy and its impact on the functional distribution of real income in the United States: A general equilibrium analysis.

Tercan Baysan, Ph.D. Minnesota 1974. Economic implications of Turkey's entry into the Common Market.

Arlene E. Bearman, Ph.D. Oklahoma. Eurodollar interest rates.

Michel Bergeron, Ph.D. Western Ontario 1974. Generalization of the production side of the trade model.

Robert R. Blake, Ph.D. Michigan 1974. Import controls in Tunisia.

Robert C. Bradshaw, Ph.D. Maryland 1973. A theoretical and empirical analysis of *U.S.* bank Eurodollar borrowing, 1966–1971.

William H. B. Brittain, Ph.D. Chicago 1975. Monetary factors in the French balance of payments: 1880–1913.

Ehsan Choudhri, Ph.D. Chicago 1974. The direction of trade in individual manufactured products: An analysis of the influence of factor proportions, economies of scale, and product innovations.

Kevin M. Cleaver, Ph.D. Fletcher School 1975. The effective rate of subsidy and resource allocation in Zaire.

Dean De Rosa, Ph.D. Oregon 1975. A survey of recent world trade models with an application to the 1971 world currency realignments.

Walter Enders, Ph.D. Columbia 1974. A two-country portfolio balance model.

Rodney E. Falvey, Ph.D. Rochester 1975. A comparison of the effects of price and quantitative restrictions on trade and investment.

Richard T. Freeman, Ph.D. Stanford 1975. Technology differences and patterns of foreign trade and investment.

Shukri M. Ghanem, Ph.D. Fletcher School 1975. The pricing of Libyan crude oil.

George Giannakouros, Ph.D. Iowa 1975. The economics of Greek shipping: Growth, problems, and alternatives.

Edward M. Graham, D.B.A. Harvard 1974. Oligopolistic imitation and European direct investment in the United States.

A. N. Hakam, Ph.D. New School 1974. Foreign industrial investments in Nigeria.

Donald P. Halstead, Ph.D. Florida State 1974. An analysis of the economic effects of free trade between Canada and the United States.

Salim E. Harik, Ph.D. Wayne State 1975. The effects of foreign investment controls on the *U.S.* balance of payments: Recoupment periods and internal discount rates.

William L. Hemphill, Ph.D. Princeton 1975. The effect of foreign exchange receipts on imports of less developed countries.

Gary K. Hewitt, Ph.D. Yale 1975. *U.S.* penetration of Canadian manufacturing industries.

John R. Hewson, Ph.D. Johns Hopkins 1975. Liquidity creation and distribution in Euro-currency markets.

John Hoftyzer, Ph.D. Indiana 1974. Empirical testing of some of the modern theories of international trade.

Peter Hooper III, Ph.D. Michigan 1975. An analysis of aggregation error in *U.S.* merchandise trade equations 1954 I—1972 IV.

William K. Hutchinson, Ph.D. Iowa 1975. Foreign exports and regional growth: The United States, 1870–1910.

Usameh F. Jamali, Ph.D. Fletcher School 1975. The energy policy of the European coal and steel community, attitudes and responses of the high authority.

DeAnne S. Julius, Ph.D. California (Davis) 1975. A monetarist model of the demand for international reserves.

Akbar Karbassian, Ph.D. New Mexico 1975. The conventional and the evolutionary doctrine of capital movements and technological diffusion in the process of economic development.

William Kleiner, Ph.D. Michigan State 1974. Balance-of-payments adjustment policies in Finland, Denmark, and Norway.

Shoichi Kobayashi, Ph.D. Hawaii 1975. Reexamination of the Hecksher-Ohlin theorem and the Leontief paradox under variable capital utilization.

Moon H. Lee, Ph.D. Chicago 1974. Excess inflation and currency depreciation.

Tai-Young Lee, Ph.D. Indiana 1974. The trade relation between Japan and Korea: An empirical study of the Heckscher-Ohlin theorem of comparative advantage.

Robert F. Lucas, Ph.D. Brown 1975. Monetary union: An alternative to free exchange rates?

David S. McClain, Ph.D. Massachusetts Institute of Technology 1974. Foreign investment in *U.S.* manufacturing and the theory of direct investment.

Eugenio A. Maffucci, Ph.D. Minnesota 1974. Export earnings instability—the Argentine case.

Daryosh Malakpour, Ph.D. Princeton 1975. Determinants of the commodity composition of trade among nations: The factor-endowment hypothesis reexamined.

Jorge Marquez-Ruarte, Ph.D. Chicago 1975. The two-sector model of production.

John E. Marthinsen, Ph.D. Connecticut 1975. European monetary unification and its effect upon the Eurocurrency market.

Angelo Mascaro, Ph.D. Michigan 1974. An analysis of limited flexibility policies for balance-of-payments adjustment.

Ricardo Mejias, Ph.D. Florida State 1974. The effect of economic growth on the trade balance, the exchange rate, and relative prices.

Anne-Marie Meulendyke, Ph.D. Chicago 1975. Causes and consequences of the Eurodollar expansion.

Dennis R. Murphy, Ph.D. Indiana 1974. The internal-external adjustment problem with labor market distortions.

Samuel C. Nana-Sinkam, Ph.D. George Washington 1975. Monetary obstacles to the expansion of intra-African trade and the development of monetary cooperation in Africa: Application of a theory of optimum currency areas in Africa.

Anyegbunam William Obi, Ph.D. Michigan State 1974. Export instability, protection-induced cost-price distortions, and economic growth in less developed countries.

John Ohly, Jr., Ph.D. Boston 1975. Some effects of export policy on the distribution of income in post-World War II Argentina.

Hiroshi Ono, Ph.D. Brown 1975. Long-run capital movements, tariffs, and terms of trade.

Thomas M. Patrick, Ph.D. Kentucky 1974. Mature creditorship and the *U.S.* balance of payments.

Harold Payson III, Ph.D. Fletcher School 1975. Forward exchange, arbitrage, and the financial cost of forward intervention.

Bruce L. Petersen, Ph.D. Georgetown 1975. Distributional effects of changes in domestic tax structure in an open economy with fixed factor supplies.

Jose M. Pinera, Ph.D. Harvard 1975. World income redistribution through trade.

Peter K. Pollak, Ph.D. Minnesota 1974. Economic analysis of oilseed markets in Thailand.

John B. Ridpath, Ph.D. Virginia 1975. An econometric and historical study of the freely floating Canadian dollar, 1931–39.

John Roemer, Ph.D. California (Berkeley) 1975. United States-Japanese competition: A study of international trade and investment rivalry.

PAUL RUCKMAN, Ph.D. Kansas 1975. An analysis of the effects of the common agricultural policy on productivity and wages in the European economic community.

ALAN M. RUGMAN, Ph.D. Simon Fraser 1974. Foreign operations and the stability of *U.S.* corporate earnings: risk reduction by international diversification.

WILLIAM T. RULE, Ph.D. Virginia Polytechnic Institute 1975. Money, income, and prices: The velocity of money in nine major *OECD* economics.

GHASSEM SALEHKHOU, Ph.D. New School 1974. Commercialization of technology in developing countries: Transfer of pharmaceutical technology to Iran.

RICHARD E. SCHATZ, Ph.D. Hawaii 1974. Optimal holdings of international reserves: The case of Thailand.

EKKEHARD F. STILLER, Ph.D. Hawaii 1975. The Eurobond market and interest rate linkages.

ROLAND H. E. STILLER, Ph.D. Hawaii 1975. German official foreign reserve holdings 1970–1973, optimal or nonoptimal.

CAROLYN C. STREMLAU, Ph.D. Fletcher School 1975. Indigenization of management in foreign-owned firms in Nigeria.

ALFREDO S. TEPPERBERG, Ph.D. Cincinnati 1975. Trade creation and trade diversion in the Latin American free trade area.

NGUYEN THI BICH THUY, Ph.D. Michigan State 1974. An application of optimal control theory to policies for internal and external balance.

EUGENE YU-SENG TSOA, Ph.D. Notre Dame 1975. A study on the actual deviations from the interest parity theory.

HERMAN VERWILST, Ph.D. Johns Hopkins 1975. The foreign exchange market with a dual exchange rate mechanism.

DALE VORDERLANDWEHR, Ph.D. Wayne State 1974. Toward an integration of international monetary economics.

RUPERT D. WORRELL, Ph.D. McGill 1975. Central bank balance of payments in the Caribbean.

RAM P. YADAV, Ph.D. Cornell 1974. An econometric model for the foreign trade sector of India: 1960–61—1971–72.

YUNG YONG YANG, Ph.D. Oregon 1974. The response of Korean exports to variations in the effective export price.

Business Administration; including Business Finance and Investment, Insurance, Marketing, and Accounting

DAVID M. AHLERS, Ph.D. Carnegie-Mellon 1975. An evaluation of alternative investment decision-making systems.

WILLIAM R. ALLEN, Ph.D. Florida 1975. A comparative analysis of black and white leadership in a naturalistic setting.

JASPER H. ARNOLD III, D.B.A. Harvard 1975. A study of privately placed incentive financings.

NORMAN BAGLINI, Ph.D. Temple 1974. Risk management in American multinational international corporation.

C. LANCE BARNETT, Ph.D. Oregon 1975. Underwriter pricing decisions in the tax exempt bond market and the principle of safety first.

WILLIAM O. BEARDEN, Ph.D. South Carolina 1975. Situational and attitude models of consumer brand choice for soft drinks and beer.

JESSE V. BOYLES III, Ph.D. Florida 1975. Accounting for long-lived productive resources: The development and evaluation of general energy systems measurement procedures.

STEWART L. BROWN, Ph.D. Florida 1974. The adjustment of stock prices to new information: A test of market efficiency.

CHERYL A. CASPER, Ph.D. Case Western 1975. Pricing models and informational requirements for library services.

B. H. CORUM, Ph.D. Florida 1975. A comparative and programmatic approach to organizational control, and its application to the hospital.

GERALD EYRICH, Ph.D. Claremont 1975. A risk return analysis of corporate financial behavior.

CHARLES L. HEDRICK, Ph.D. Carnegie-Mellon 1975. A computer program to learn production systems using a semantic net.

DANNY R. HINES, Ph.D. Florida 1975. Adapting the flow of resources concept to the business enterprise.

HARRISON E. HUNSICKER, Ph.D. Colorado State 1975. An operational analysis of the RRPM: A case study at CSU.

MICHAEL KAVANAUGH, Ph.D. Cincinnati 1975. The effect of market and organization characteristics on the relative size of nonproduction worker employment.

DAVID B. KISER, D.B.A. Harvard 1975. Allocating resources for consumer projects: The corporate dilemma.

PARAMESWAR KRISHNAKUMAR, Ph.D. Florida 1974. An exploratory study on the influence of origin on the product images of persons from selected countries.

WILLIAM C. LATHAM, Ph.D. South Carolina 1975. The effect of inconsistency on forecasting earnings per share trend.

MAWSEN LIAO, Ph.D. Florida 1974. A comparison of stochastic linear programming with mean value linear programming for production and profit planning under conditions of uncertainty.

THOMAS R. LLOYD, D.B.A. Harvard 1975. Product innovation: The correspondence of organization and decision process to environment and strategy.

WILLIAM A. MCEACHERN, Ph.D. Virginia 1975. Management incentives and firm performance under three different control conditions.

RODNEY P. MALONE, Ph.D. Florida 1974. Automatic dividend reinvestment plans of nonfinancial corporations.

Osama I. Mikhail, Ph.D. Carnegie-Mellon 1975. On the economics and design of computer networks for distributed data processing.

Linda C. Mitchusson, Ph.D. Arkansas 1975. The relationship of executive compensation to the profitability of management-controlled firms.

Richard W. Mizerski, Ph.D. Florida 1974. Attribution theory and consumer processing of unfavorable information about products.

Michael J. Murray, Ph.D. Notre Dame 1975. The cost of capital: A test of alternative views.

Kotaro Narizoe, Ph.D. Claremont 1975. Selection of optimum bond portfolio.

Jerald L. Nelson, Ph.D. North Carolina State 1974. Taxation and other economic effects on the capital structure of farm supply cooperatives.

Rodolfo E. Paiz, D.B.A. Harvard 1974. Experimenting with new concepts of retail good distribution in a developing environment.

Roger C. Poole, Ph.D. South Carolina 1975. Blue sky laws and the registration of new issues of common stock: An empirical study.

Coimbatore K. Prahalad, D.B.A. Harvard 1975. The strategic process in a multinational corporation.

Roland R. Reitter, D.B.A. Harvard 1975. Georges Pompidou—the growth and vicissitudes of power through consensus.

James H. Scott, Jr., Ph.D. Carnegie-Mellon 1975. Firm valuation, optimal capital structure, and the risk of bankruptcy.

Andrew J. Silver, D.B.A. Harvard 1975. The management of creative process in a temporary system.

Bruce A. Smith, Ph.D. Virginia Polytechnic Institute 1975. One possible approach to pricing at a university computer center.

John H. Spitzer, Ph.D. Duke 1975. The rates of return to capital, advertising, and research and development: Their sensitivity to the structure of the distributed time lag.

Robert Taggart, Ph.D. Massachusetts Institute of Technology 1974. A model of corporate financing decisions.

Tharsi Taillieu, Ph.D. Carnegie-Mellon 1975. Some effects of reward systems and feedback methods on individual decisions to allocate group resources.

John C. Warner, Ph.D. West Virginia 1974. Long-term interest rate expectations and short-run changes in business fixed investment.

James K. Weeks, Ph.D. South Carolina 1974. A simulation study of due-date assignment policies in a dual-constrained job shop.

Peter U. Welam, Ph.D. Carnegie-Mellon 1975. Mathematical programming models for marketing and production management.

Joesph T. Williams, Ph.D. Harvard 1974. Essays in portfolio theory.

Industrial Organization and Public Policy; including Economics of Technological Change, and Industry Studies

John M. Albertine, Ph.D. Virginia 1975. The behavior of broad spectrum antibiotics prices 1951–60: An economic analysis.

Benjamin J. Allen, Ph.D. Illinois 1974. The economic effects of rail abandonment on communities: A case study.

George A. Aragon, D.B.A. Harvard 1975. The impact of price controls on large manufacturing firms: An analysis of the base period profit margin limitation of Phase II.

Thomas A. Bankston, Ph.D. Florida 1975. The relationship between market price and book value for regulated utilities.

Lawrence W. Bates, Ph.D. Texas (Austin) 1975. The petroleum industry in Brazil.

Mary F. Berglund, Ph.D. Nebraska 1974. An economic evaluation of the system of railroad freight car supply in the United States.

Linda T. Bilheimer, Ph.D. Harvard 1974. Hospitals in Arkansas: A study in firm and industry behavior.

Frank V. Billes, Ph.D. California (Los Angeles) 1974. An econometric study of investment behavior in the electric power industry.

Joan Bodoff, Ph.D. New School 1974. Monopoly and price revisited.

Lawrence J. Boyle, Ph.D. Maryland 1973. The development of the Quebec hydroelectric industry.

Paul S. Brandon, Ph.D. Harvard 1974. The electric side of combination gas and electric utilities.

Yale M. Braunstein, Ph.D. Stanford 1975. Dynamics of inter- and intramedia competition: Effects of technological and demographic change.

Luvonia J. Casperson, Ph.D. Oklahoma. The economic impact of air quality regulations on the electric power and coal industries in Oklahoma.

John Cirace, Ph.D. Columbia 1975. The law and economics of locationally exclusive and territorially restricted franchises.

Philip L. Cline, Ph.D. Oklahoma State 1975. Sources of productivity change in *U.S.* agriculture.

John H. Crockett, Jr. Ph.D. North Carolina 1975. An analysis of differential pricing of electric power to residential users.

S. Michael Dean, D.B.A. Harvard 1974. The multinational corporation and third world economic nationalism: Organizational response to environmental crisis.

James B. Delaney, Ph.D. Boston College 1974. On the measurement of concentration: A cross-sectional analysis of alternative indicies.

Allen E. Early, Jr., Ph.D. Oklahoma. Regulations of life insurance in Texas: A study of the impact of regulatory policies on industry structure, behavior on performance.

Richard R. Edwards, Ph.D. North Carolina State 1974. Estimating the demand for cable television.

Avrohn Eisenstein, Ph.D. New School 1975. Nonemployer retail trade establishments in the United States, 1939–67.

Aly Alp Ercelawn, Ph.D. Vanderbilt 1974. An econometric analysis of the residential demand for electricity in the United States.

David J. Faurot, Ph.D. Northwestern 1975. An exact flexible accelerator model of input demand.

Gary Fauth, Ph.D. Harvard 1974. Bus priority schemes and economic efficiency.

John R. Finlay, Ph.D. Queen's 1974. Structural and dynamic change in grocery retailing in Canada, 1930–1970.

John A. Fox, Ph.D. Brown 1975. An empirical test of the Averch-Johnson hypothesis in the electric utility industry.

Pierre Gadonneix, D.B.A. Harvard 1974. The *plan calcul:* An attempt to meet the *U.S.* challenge in the French computer industry.

Theodore A. Gebhard, Ph.D. Illinois 1975. A pricing-output optimization model for a professional team sports league.

Anthony George, Jr., Ph.D. Virginia 1974. Second best pricing and the postal service.

Frank M. Gollop, Ph.D. Harvard 1974. Modelling technical change and market imperfections: An econometric analysis of *U.S.* manufacturing, 1947–71.

Leonard M. Good, Ph.D. Western Ontario 1974. An econometric model of the Canadian cable television industry and the effects of CRTC regulation.

James G. Halteman, Ph.D. Pennsylvania State 1974. Should municipal authority water firms be regulated? A comparative study between private and municipal authority firms.

John R. Haring, Ph.D. Yale 1975. Competition, regulation, and performance in the commercial radio broadcasting industry.

Mel Horwitch, D.B.A. Harvard 1975. Technological evaluation and the rise of a societal issue: A history of sonic boom decision making in the US/SST program (1964–69).

Keith M. Howe, Ph.D. Nebraska 1974. The impact of the economic and regulatory setting on the public utility's cost of capital.

Frederic Y. Jenny, Ph.D. Harvard 1975. Industrial structure and economic performance in France.

Frank P. Johnson, Ph.D. Colorado State 1975. Differential tuition determined by unit cost for higher education.

Marc A. Johnson, Ph.D. Michigan State 1975. Market and social investment and disinvestment in railroad branch lines: Evaluation procedures and decision criteria.

William E. Kamps, Ph.D. Washington State 1974. The score of "the market": Empirically, theoretically, and legally.

Daniel P. Kaplan, Ph.D. Michigan 1974. Growth in demand, technological change, and changing market shares.

John M. Karber III, Ph.D. Arkansas 1975. An analysis of structure and conduct in American industry.

Kevin R. Kelly, Ph.D. Oregon 1974. An empirical analysis of individual and community demand for urban transportation.

Michael P. Kennedy, Ph.D. Harvard 1974. An economic model of the world oil market.

Thomas E. Kennedy, Ph.D. Johns Hopkins 1975. Regulatory control of monopoly: Limitations, methods, and effects.

Ezra Kessel, Ph.D. New School 1974. Economics of the refrigerator and other appliance industries.

Ramin Khadem, Ph.D. McGill 1975. An econometric model of the demand for international voice telecommunication from Canada.

John W. Kostolansky, Ph.D. Columbia 1975. An economic analysis of the United Shoe Machinery case.

Robert Kudrle, Ph.D. Harvard 1974. Agricultural tractors: A world industry study.

Nathan N. Kuma, Ph.D. Boston 1975. Import-substitution and industrialization in Ghana: A case study of the Ghanaian sugar industry.

Rodger S. Lawson, Ph.D. Utah 1975. Benefit-cost analysis and the allocation of biomedical research resources.

Alexander Lebanon, Ph.D. Carnegie-Mellon 1975. The estimation of cross-sectional dependence and a cross-sectional time-series study of the demand for energy and fuels in OECD countries.

Richard Levin, Ph.D. Yale 1974. Technological change, economies of scale, and market structure.

Thomas D. McKinnon, Ph.D. Arizona 1975. Prices, qualities, and market structure in the high fidelity component industry.

Michael McPherson, Ph.D. Chicago 1974. The effects of public on private college enrollment.

Aroop K. Mahanty, Ph.D. Colorado State 1975. Behavior of the firm under a regulatory constraint.

Robert Maisel, Ph.D. California (Berkeley) 1975. Structural change and the rate of return in *U.S.* manufacturing industries, 1954–63.

Howard Marvel, Ph.D. Chicago 1975. Determinants of the behavior of retail gasoline prices: An application of the economics of information.

Patricio Meller, Ph.D. California (Berkeley) 1975. Production functions and efficiency frontiers for industrial establishments of different sizes: The Chilean case, year 1967.

Gary M. Mennem, Ph.D. Kansas State 1974. Effects of railroad abandonment on grain producers and grain elevator supply areas in north central Kansas.

Lynn O. Michaelis, Ph.D. Washington State. Trade credit, firm size, and market power in relation to monetary and industrial organization theory.

RONALD H. MILLEN, Ph.D. Concordia 1974. The use of a Bell Canada production model in an automatic rate adjustment mechanism.

JAMES F. MONGOVEN, Ph.D. Notre Dame 1975. Competition in the soft drink industry.

GERALD R. MOODY, Ph.D. George Washington 1975. Entry by foreign firms: The case of the heavy electrical equipment industries.

ARTHUR J. MOSES, Ph.D. Duke 1975. Projections of the energy base of the Soviet transport sector.

ROBERT A. MULLER, Ph.D. Toronto 1975. A simulation of adjustment to pollution control costs in the pulp and paper industry.

FRANCIS P. MULVEY, Ph.D. Washington State 1974. The economic future of Amtrack.

MICHAEL M. MURPHY, Ph.D. California (Los Angeles) 1974. A theory of regulatory behavior.

CHARLES W. NEEDY, Ph.D. Florida 1974. Regulation-induced distortions: A geometric integration.

PAUL A. NELSON, Ph.D. Wisconsin (Madison) 1974. The effect of market structure on efficiency in the electric power industry.

JAMES J. O'CONNELL, Ph.D. Fordham 1975. An analysis of the economies of scale resulting from regional approaches to solid waste management.

SHARON M. OSTER, Ph.D. Harvard 1974. The incidence of local government expenditures on water pollution abatement.

RAYMOND M. OWENSBY, Ph.D. Kansas State 1974. The utilization of railroad wheat cars.

JOHN C. PANZAR, Ph.D. Stanford 1975. Regulation, service quality, and market performance: A model of airline rivalry.

MARIO PICCONI, Ph.D. Rutgers 1975. A microeconometric investigation of the firm's advertising decision.

THOMAS A. PINFOLD, Ph.D. Minnesota 1975. The modal split problem in urban transportation.

GEORGE POLAK, Ph.D. Case Western 1974. Economic effects of technological innovations in the shallow-draft water transport industry.

DONALD W. RAMEY, Ph.D. Washington State 1975. The determinants of research effort for industry leaders in manufacturing.

FRANCIS RENAUD, Ph.D. New School 1975. The New York state commercial banking industry: An econometric analysis of the relationship between performance, market share, and efficiency.

BURTON G. ROTHBERG, D.B.A. Harvard 1975. A decision theoretic model of the eastern hemisphere oil exploration.

CALVIN T. ROUSH, Ph.D. California (Santa Barbara) 1974. The California crude oil market: An examination of the posted price structure.

JOHN G. ROWSE, Ph.D. Minnesota 1974. Toward optimal capacity expansion for an electric ultility: A mixed integer programming approach.

LESLIE S. SCRUGGS, JR., Ph.D. Vanderbilt 1974. Economies of scale in securities-market information.

FLORENCE SETZER, Ph.D. Yale 1974. Technological change over the life of a product: Changes in skill inputs and production processes.

ROBERT D. SHRINER, Ph.D. Indiana 1974. An econometric analysis of the demand and supply for scrap iron and steel.

STEPHEN D. SILBERMAN, Ph.D. Yale 1975. The market, the firm, and delivery of legal services.

RICHARD H. SINES, Ph.D. Ohio State 1975. A cross-sectional micro analysis of production: The case of Venezuelan manufacturing.

BRUCE A. SMITH, Ph.D. Indiana 1974. Technological innovation in electric power generation, 1950–70.

BRUCE R. SNAPP, Ph.D. Michigan 1974. Market structure, firm size, and merger strategies.

MARCELLUS S. SNOW, Ph.D. California (Berkeley) 1975. Cost structure and charging policy in international commercial satellite communications.

HARRY J. STORY, Ph.D. Oregon 1975. An economic analysis of the plywood futures market.

JOHN P. TIEMSTRA, Ph.D. Massachusetts Institute of Technology 1975. Protest through market action.

ROY G. VAN TIL, Ph.D. Boston College 1975. The content of *U.S.* automobile advertising 1949–73: Impact on the shaping of wants and product development.

WILLIAM J. VAUGHAN, Ph.D. Georgetown 1975. A residuals management model of the iron and steel industry: A linear programming approach.

ANTHONY J. VIGNOLA, JR., Ph.D. Ohio State 1975. The taxation of regulated industries: An analysis of selected economic issues relevant to Ohio.

KENNETH WARNER, Ph.D. Yale 1974. The diffusion of leukemia chemotherapy: A study in the nonmarket economics of medical care.

FREDERICK R. WARREN-BOULTON, Ph.D. Princeton 1975. Vertical control with variable proportions.

PAULINE M. WEBER, Ph.D. Oklahoma State 1974. Risk and advertising in an oligopoly game.

JONATHAN B. WELCH, Ph.D. Connecticut 1975. Telecommunications: Competition and technological change.

SUSAN M. WILCOX, Ph.D. California (Santa Barbara) 1975. Joint venture bidding and entry in the market for offshore petroleum leases.

ROBERT W. WILSON, Ph.D. Yale 1975. The sale of technology through licensing.

ALLAN H. ZELENITZ, Ph.D. Princeton 1975. The Ford Autolite divestiture case: An analysis of the attempted promotion of competition in related goods markets.

Agriculture and Natural Resources

LAWRENCE W. ABRAMS, Ph.D. Washington (St. Louis) 1974. Operational use of corrective taxes for pollution control.

DAVID R. ALLARDICE, Ph.D. Colorado State 1975. Economic and legal protection for aquatic habitats.

STEPHEN O. ANDERSEN, Ph.D. California (Berkeley) 1974. Economics of nuclear power plant location with emphasis on the coastal zone.

CHARLES B. AUSTIN, JR., Ph.D. Clark 1974. The economic implications of overfishing in a multispecies fishery.

ANDRÉS BAJUK, Ph.D. California (Berkeley) 1975. A model of the distribution of income from technological change in agriculture.

OM BATISH, Ph.D. Claremont 1975. Effects of the agricultural land taxation in the post-independence Punjab.

HENRY BEALE, Ph.D. Chicago 1974. The use of capitalized values of land and water rights in evaluating irrigation projects.

FREDERICK M. BLANK, Ph.D. Duke 1975. An economic theory of pollution.

ROE E. BORSDORF, Ph.D. Kansas State 1975. An investigation into forecasting planted crop acreages in the short run as a quantitative measurement tool for agribusiness management.

ABDALLAH RASHID BU-HABIB, Ph.D. Vanderbilt 1975. The long-run supply of new reserves of crude oil in the United States, 1966–73.

CARL CHI-JIAU CHEN, Ph.D. Iowa State 1975. Quadratic programming models of *U.S.* agriculture in 1980: With alternative levels of grain exports.

LEE CHRISTENSEN, Ph.D. Michigan State 1975. An economic and institutional analysis of land treatment as a wastewater management alternative for southeastern Michigan.

DONALD. V. DEBORD, Ph.D. North Carolina State 1974. Demand for and cost of salt marsh mosquito abatement.

JAMES F. DEEGAN, Ph.D. Southern Methodist 1975. An econometric model of the Gulf Coast oil and gas producing basin.

LUIS E. DI MARCO, Ph.D. California (Berkeley) 1974. The regional income distribution in Argentina.

ABDULAZIZ M. DUKHEIL, Ph.D. Indiana 1974. An optimum base for pricing middle eastern crude oil.

GEORGE A. ECKER, Ph.D. North Carolina State 1974. Optimal replacement pattern of standard apple trees with dwarfs.

DAVID L. FORESTER, Ph.D. Michigan State 1974. The effects of selected water pollution control rules on the simulated behavior of beef feedlots, 1974–85.

RICHARD G. FRITZ, Ph.D. Georgetown 1975. A public investment scheme for water resource management in a rapidly growing community.

ROBERT F. GLENN, Ph.D. Missouri 1975. A farm point pricing system for fluid milk in selected midwestern markets.

FREDERICK W. GRAMLICH, Ph.D. Harvard 1975. Estimation of net benefits of improvements in Charles River water quality.

ALLEN GROMMET, Ph.D. Michigan State 1975. The influence and behavior of interest group leaders in the public decision process: A model with four rural applications.

BARBARA HERZ, Ph.D. Yale 1974. Demographic pressure and economic change: The case of Kenyan land reforms.

HENRY W. HERZOG, JR., Ph.D. Maryland 1974. The economics of regional water quality management (A case study of river water quality in the Chesapeake Bay region).

DARRELL L. HUETH, Ph.D. California (Berkeley) 1974. Optimal agricultural pest management under the condition of increasing pest resistance with an application to the cereal leaf beetle.

RAPHAEL U. IGWEBUIKE, Ph.D. Stanford 1975. Barriers to agricultural development: A study of the economics of agriculture in Abakaliki area, Nigeria.

BRUCE B. JOHNSON, Ph.D. Michigan State 1975. Farmland leasing as a means of resource control in *U.S.* land-based agriculture.

F. REED JOHNSON, Ph.D. State University of New York (Stony Brook) 1974. Income distributional effects of environmental policy: An application to sulphur dioxide pollution.

MUHAMMAD JAMEEL KHAN, Ph.D. Colorado State 1975. Economics of farm mechanization and water development policies in Pakistan.

FRED KING, Ph.D. Columbia 1975. An economic analysis of small farm behavior.

ROBERT J. KORBACH, Ph.D. Maryland 1973. A regional environmental input-output model.

THOMAS V. KOSZAREK, Ph.D. Minnesota 1974. Market performance of Minnesota retail farm supply cooperatives.

YUNG C. LEE, Ph.D. Michigan State 1975. Adjustment in the utilization of agricultural land in south central Michigan with special emphasis on cash-grain farms.

TRACY R. LEWIS, Ph.D. California (San Diego) 1975. Optimal resource management under conditions of uncertainty: The case of an ocean fishery.

DENNIS R. LIFFERTH, Ph.D. Iowa State 1974. An economic analysis of alternative rail-based grain distribution systems.

MICHAEL LOPEZ, Ph.D. Yale 1975. The determinants of income and its distribution in four villages in India.

JOHN MCKEON, Ph.D. Michigan State 1975. Farm commodity programs: Their effect on plantings of feed grains and soybeans.

JAMES J. MCRAE, Ph.D. Western Ontario 1974. Optimal natural resource exploitation by open economies.

MARVIN B. MARTIN, Ph.D. Iowa State 1974. An economic analysis of quality and grading in corn marketing.

JOHN J. MEEK, Ph.D. Michigan State 1975. Simulation of the cattle-calves subsector in a developed economy with special reference to the Canadian cattle herd.

KENNETH M. MENZ, Ph.D. Minnesota 1974. The impact of mobile processing plants on the production and distribution of frozen peas.

GENE MILLER, Ph.D. Missouri 1974. An economic analysis of basic grains in El Salvador.

JOHN R. MILLER, Ph.D. Washington (St. Louis) 1974. Economics of agricultural pest control: Productivity, pesticide and the prisoner's dilemma.

JOHN A. MIRANOWSKI, Ph.D. Harvard 1975. The demand for agricultural crop chemicals under alternative farm program and pollution control solutions.

SHEILA J. MOORE, Ph.D. Arizona 1975. Externality considerations and their effect on the supply of local government services.

VISHNUPRASAD SSV NAGADEVARA, Ph.D. Iowa State 1974. A modeling approach to the economic impacts of local environmental controls in Iowa and the rest of the country.

MANH H. NGUYEN, Ph.D. Toronto 1975. Essays on the optimal dynamic exploitation of natural resources and the social rate of discount.

WILLIAM C. NICKEL, Ph.D. California (Berkeley) 1975. Efficiency in resource use in paddy-intensive south Indian agriculture.

JAMES E. NICKUM, Ph.D. California (Berkeley) 1975. A collective approach to water resource development: The Chinese commune system, 1962–72.

CLAYTON W. OGG, Ph.D. Minnesota 1974. Sources of agricultural productivity differences in North America.

ANGELOS PAGOULATOS, Ph.D. Iowa State 1975. Major determinants affecting the demand and supply of energy resources of the United States.

SUKRAJ S. PANNU, Ph.D. Colorado 1975. Optimal exploitation of a renewable natural resource under value maximization.

BURT E. POWELL, Ph.D. Duke 1975. Land tenure on northern plains Indian reservations.

GEORGE PROVENZANO, Ph.D. Illinois 1974. The economics of environmental quality management: A general systems approach.

POROOR RADHAKRISHNAN, Ph.D. Stanford 1974. The role of rubber in the West Malaysian economy.

RADZUAN A. RAHMAN, Ph.D. Cornell 1974. The economics of the timber industry in West Malaysia: A case study.

PHILIP L. RAIKES, Ph.D. Stanford 1975. The development of mechanized commercial wheat farming in North Iraqw, Tanzania.

GAURISANKAR RAYCHAUDHURI, Ph.D. Stanford 1975. The scope and function of commodity futures in India (1954–70).

SHYAMAL ROY CHOWDHURY, Ph.D. Iowa State 1974. Effects of rural industrialization on rural development in Iowa.

MUSA AHMAD SAMMAN, Ph.D. Missouri 1974. Analysis of factors influencing loan losses of production credit associations.

STEVEN T. SONKA, Ph.D. Iowa State 1974. Alternative futures of exports, farm-size, and government programs for American agriculture.

DANIEL R. STRANG, Ph.D. Cornell 1975. An economic analysis of the sources and magnitudes of inefficiency in bulk milk assembly in New York State.

PAUL C. TAYLOR, Ph.D. Kansas 1975. The economics of wildlife management.

PETER E. TEMU, Ph.D. Stanford 1974. Marketing board pricing and storage policy with particular reference to maize in Tanzania.

RONALD L. THOMPSON, Ph.D. Michigan State 1975. Description and analysis of limited resource farmers in Michigan.

DANIEL S. TILLEY, Ph.D. Iowa State 1974. Hog producers' marketing decisions.

STEPHEN M. TUGWELL, Ph.D. Queen's 1974. Production and efficiency in the Maritimes' lobster fishery.

TERRENCE S. VEEMAN, Ph.D. California (Berkeley) 1975. Economic consequences and policy implications of the Green Revolution in India with particular emphasis on water resources policy in Punjab.

AR-YOUNG WANG, Ph.D. Cincinnati 1975. Dynamic analysis of natural resource economics in fishing.

CLEMENT E. WARD, Ph.D. Kansas State 1974. A decision model applied to forward sale slaughter hog contracts.

JAMES N. WETZEL, Ph.D. North Carolina 1974. A distance congestion model for evaluating recreation resources.

CHESTER E. YOUNG, JR., Ph.D. North Carolina State 1974. Economic analysis of municipal waste water treatment systems.

Manpower, Labor, and Population; including Trade Unions and Collective Bargaining

NAIM AHMED, Ph.D. South Carolina 1974. The choice of contraceptive technology by a family planning clinic in a developing country: An analysis of the economic and noneconomic constraints.

CONSTANTINE C. AZARIADIS, Ph.D. Carnegie-Mellon 1975. A theory of labor contracts.

WILLIE H. BAILEY, Ph.D. Illinois 1974. Economic analysis of regional wage adjustments.

NANCY S. BARRY, Ph.D. Northwestern 1975. The international migration patterns of professional manpower: The case of British economists.

ANDREA BELLER, Ph.D. Columbia 1975. The effect of Title VII of the Civil Rights Act of 1964 on the economic position of black males.

FRANCINE D. BLAU, Ph.D. Harvard 1975. Pay differentials and differences in distribution of employment of male and female office workers in northeastern cities.

BARRY A. BLUESTONE, Ph.D. Michigan 1974. The personal earnings distribution: Individual and institutional determinants.

GEORGE J. BORJAS, Ph.D. Columbia 1975. Job investment, labor mobility and the earnings of older men.

MICHAEL L. BROOKSHIRE, Ph.D. Tennessee 1975. Collective bargaining in the Tennessee Valley Authority: The salary policy employee experience.

CHARLES C. BROWN, Ph.D. Harvard 1974. Young men: Human capital and earnings.

GARY D. BROWN, Ph.D. Illinois 1974. Male-female wage differentials.

DENNIS M. BYRNE, Ph.D. Notre Dame 1975. Analysis of time allocation patterns of inner city blacks.

JOHN M. CAMPBELL, JR., Ph.D. Oklahoma. Household demand: A synthesis of interdisciplinary theory and empiricism.

EMMETT E. CARADINE, JR., Ph.D. Arkansas 1975. Court administration of a workmen's compensation system: A study of the Louisiana experience.

ALBERT M. CEDERLUND, Ph.D. Clark 1974. The effect of pension plans on quit rates in the Worcester industrial area.

LUKE KIN-PONG CHAN, Ph.D. California (Berkeley) 1975. Nonpecuniary return to work: Theory and empirical evidence based on the value of commutation time.

PETER T. CHINLOY, Ph.D. Harvard 1974. Issues in the measurement of labor input.

GARLAND R. CHRISTOPHER, Ph.D. California (Berkeley) 1975. The effects of social background on regionally mobile and nonmobile men: A case study.

SONIA R. CONLY, Ph.D. South Carolina 1975. Subsidized day care and the employment of lower income mothers: A case study.

MARIO CORBO, Ph.D. Chicago 1974. Schooling, experience, and wages in Santiago, Chile.

DANIEL Z. CZAMANSKI, Ph.D. Syracuse 1975. A study of factors affecting the productivity of spending on municipal preventive services: A case study of fire departments.

HENRY B. DARLING, JR., Ph.D. Arkansas 1975. An investigation of the potential benefits of a productive education industry in an underdeveloped area and the explicit costs of a maximum subsidy or bill of rights for continuing education for all public school teachers and supervisors of the SMSA concerned.

ALBERTO DE MELLO E SOUZA, Ph.D. Michigan 1975. Rates of return, occupational mobility and the labor market for industrial workers: A study of two Brazilian states.

ARTHUR G. DOBBELAERE, JR., Ph.D. Notre Dame 1975. The effect of past practice on the arbitration of labor disputes.

WILLIAM S. DONNELLY III, Ph.D. Maryland 1973. An econometric model of net labor migration in the United States, 1965.

GREG J. DUNCAN, Ph.D. Michigan 1974. Nonpecuniary work rewards: Implications for studies of earnings functions, discrimination, and labor union effects.

JAMES E. EASLEY, JR., Ph.D. North Carolina State 1974. Expenditures per child in an economic model of fertility.

JAMES H. EBLEN, Ph.D. Tennessee 1974. A study of outmigration of workers from southeastern metropolitan areas.

GARY EISEN, Ph.D. New School 1974. Economic determinants of fertility in American farm families.

ROBERT A. ELLSWORTH, Ph.D. Utah 1974. Regional demographic trends in antebellum America.

FREDERICK J. ENGLANDER, Ph.D. Rutgers 1975. Economic variables affecting the labor market status of blacks and women.

VALERIE M. ENGLANDER, Ph.D. Rutgers 1975. An economic analysis of the allocation of resources for rehabilitation research.

JAMES A. EVENSON, Ph.D. Colorado State 1975. Economic and statistical analysis of human capital.

DOUGLAS C. EWBANK, Ph.D. Princeton 1975. An examination of several applications of the standard pattern of age at first marriage.

CLAUDE H. FARRELL III, Ph.D. North Carolina State 1974. Forced busing and the demand for schooling.

RASHIDUR REZA FARUQEE, Ph.D. Michigan 1975. Population planning, development, and multisector programming: A dynamic model for Bangladesh.

HENRY E. FELDER, Ph.D. Stanford 1975. Job search: An empirical analysis of the search behavior of low income workers.

ROBERT S. GAY, Ph.D. Wisconsin (Madison) 1975. The impact of unions on relative real wages: New evidence on effects within industries and threat effects.

JENNIFER GERNER, Ph.D. Wisconsin (Madison) 1974. Economic implications of treating maternity leave as a temporary disability.

FLORA D. GILL, Ph.D. Stanford 1975. An analysis of Negro emigration from the southern United States: 1910–70.

RICHARD J. GOETTLE IV, Ph.D. Cincinnati 1975. Queueing approach to unemployment theory.

STEPHEN K. GRANT, Ph.D. Clark 1974. Unemployment, job vacancies, and wage rates in New Hampshire.

THOMAS HAMER, Ph.D. Claremont 1975. The labor cost of an all volunteer armed force: A comparative analysis base on police wages and occupational earnings.

MICHAEL J. HAY, Ph.D. Minnesota 1974. An economic analysis of rural-urban migration in Tunisia.

SUE E. HAYES, Ph.D. California (Berkeley) 1975. Seasonal workers in extended employment: Implications for agricultural labor policy.

BETTY C. HEIAN, Ph.D. California (Berkeley) 1975. Migration of manpower into postwar Germany from labor surplus regions of Western Europe.

STEPHEN HENDERSON, Ph.D. Wisconsin (Madison) 1975. Optimal control of a manpower hierarchy with demand uncertainty.

ROBERT S. HERREN, Ph.D. Duke 1975. Wage-price policy during the Truman Administration: A postwar problem and the search for its solution.

JAMES V. HIGGINS, Ph.D. North Carolina State 1974. An analysis of family labor supply in rural areas.

THOMAS A. HOOPENGARDNER, Ph.D. Michigan 1974. Labor migration in the Ivory Coast.

HOWARD A. HUNT, Ph.D. California (Berkeley) 1975. Registered nurse education and the registered nurse job market.

HILLARD HUNTINGTON, Ph.D. State University of New York (Binghamton) 1975. An empirical study of ethnic linkages in Kenyan rural-urban migration.

HENRY H. ITO, Ph.D. Minnesota 1974. Job search, the work-leisure tradeoff, and secondary labor force behavior.

JEROME JOFFE, Ph.D. New School 1975. Job ladders and job structures: A study of labor stratification within maintenance and repair occupations in New York City.

WILLIAM C. JOHNSON, D.B.A. Texas Tech 1975. A review and analysis of changed work schedules in hospitals.

MARVIN M. KAHN, Ph.D. Washington (St. Louis) 1974. Substitution in the labor market: A nonseparable approach.

DISMAS B. KALCIC, Ph.D. Washington (St. Louis) 1974. Interarea occupational earnings differentials: A cross-sectional analysis.

DHARAM P. KAPIL, Ph.D. California (San Diego) 1975. Empirical analysis of labor force behavior and some implications.

CONSTANTINE KAPSALIS, Ph.D. Rochester 1975. An econometric estimation of labor supply functions in Canada.

RICHARD V. KAUFFMAN, Ph.D. Colorado State 1975. A study of the educational production function.

JOHN K. KEHOE, D.B.A. Harvard 1975. Strategy in a labor union: A study of the UAW.

MASHAL KHAN, Ph.D. Stanford 1975. Socioeconomic factors in the reduction of fertility in the Islamic world with a case study of Iran.

MORRIS M. KLEINER, Ph.D. Illinois 1974. An analysis of interstate labor force migrations for manpower planning.

THOMAS J. KNIESNER, Ph.D. Ohio State 1974. Recent behavior of the "full-time" workweek in the United States.

GORDON KUBOTA, Ph.D. Claremont 1975. Technological change and labor skills substitutability: An econometric study.

EDWARD P. LAZEAR, Ph.D. Harvard 1974. The timing of technical change: An analysis of cyclical variations in technology production.

EUGENE R. LEBRENZ, D.Ed. Northern Illinois 1974. A study of large *U.S.* commercial banks to determine preferred undergraduate academic preparation for initial employment in loan officer training programs.

DEAN R. LEIMER, Ph.D. Washington (St. Louis) 1974. Optimal patterns of life cycle resource allocation with emphasis on the human capital subsector.

NICOLAOS T. LEVENTIS, Ph.D. Clark 1974. The effects of the government sector on the labor-leisure choice.

JAMES E. LONG, Ph.D. Florida State 1974. Racial discrimination in employment: A public-private sectoral analysis.

MICHAEL F. LUKOMSKI, Ph.D. Michigan State 1974. Alternatives for the training of skilled industrial labor in São Paulo, Brazil.

STEPHEN T. MARSTON, Ph.D. Michigan 1974. An econometric analysis of unemployment insurance payments: A time-series study of Detroit, Michigan.

MYLES MAXFIELD, JR., Ph.D. Maryland 1975. Studying wage changes with a simulation of the labor market.

YASH P. MEHRA, Ph.D. Minnesota 1974. Spillovers in wage determinations in *U.S.* manufacturing industries.

PATRICIO MILLAN-SMITMANS, Ph.D. Harvard 1975. The intensive use of capital in industrial plants: Multiple shifts as an economic option.

ANN V. MILLER, Ph.D. Northwestern 1974. The optimal allocation of time over the life cycle with an age-dependent utility function: Implications for the female investment in education.

GARY A. MOORE, Ph.D. Nebraska 1974. Some salary effects of professional negotiations in the public schools: The Nebraska experience.

CATHY L. MORGAN, Ph.D. South Carolina 1975. Perceived versus actual risk in occupational choice: A discriminant analysis approach.

GEORGE R. NEUMANN, Ph.D. Northwestern 1974. Wage inflation in *U.S.* manufacturing industries, 1957–69: An empirical study.

ANN E. NICHOLS, Ph.D. Colorado State 1975. The pricing of physician and lawyer services.

STANLEY D. NOLLEN, Ph.D. Chicago 1974. The supply and demand for college-educated labor.

DENNIS J. O'DONNELL, Ph.D. Pennsylvania State 1974. The microeconomics of completed family size—a simultaneous equation approach.

MARGARET OPPENHEIMER, Ph.D. Northwestern 1974. The allocation of time in economic theory and planning: An extension of input-output analysis.

HONG YOUL PARK, Ph.D. Utah State 1975. Economic determinants of rural-urban labor migration: The case of Korea.

JOHN S. PETTENGILL, Ph.D. Massachusetts Institute of Technology 1975. Labor power and wage differentials: A general equilibrium approach.

JOHN T. PICKERILL, Ph.D. Northwestern 1975. Demographic characteristics and place-to-place migration: A disaggregate approach.

DWIGHT A. PORTER, Ph.D. Maryland 1974. The four-day workweek: An examination of experience and prospects.

LEILA J. PRATT, Ph.D. Virginia Polytechnic Institute 1975. An economic analysis of occupational licensing.

GORDON D. QUICK, Ph.D. Florida 1974. Allocation of labor machine shop scheduling.

YUVAL RABINOVITCH, Ph.D. Florida 1975. Measurement of the effects of socioeconomic variables on college student employment: A test of applicable *U.S.* 1970 Census data.

RODNEY P. ROMIG, Ph.D. Nebraska 1975. Labor union growth theories and some empirical evidence: 1900–1960.

WARREN C. SANDERSON, Ph.D. Stanford 1975. Towards understanding the fertility of American women, 1920–66.

JAMES E. SAWYER, Ph.D. Utah 1975. Crucial elements in state and local manpower planning.

RONALD E. SAWYER, Ph.D. North Carolina 1974. Patterns and determinants of internal migration: The case of Mexico.

STEPHEN A. SCHNEIDER, Ph.D. Pennsylvania 1975. Competitive employment opportunities and the labor supply to a plant in a local market.

LAURENCE S. SEIDMAN, Ph.D. California (Berkeley) 1975. The design of federal employment programs: An economic analysis.

RUSSELL F. SETTLE, Ph.D. Wisconsin (Madison) 1974. The welfare economics of occupational safety and health standards.

THOMAS A. SEVERN, Ph.D. Nebraska 1974. Labor force discrimination in the west north central region: A Markovian approach.

MOHAMMAD SHADMAN-VALAVI, Ph.D. Northwestern 1974. An analysis of the determinants of migration in Iran.

HANANTO SIGIT, Ph.D. Hawaii 1975. Demographic change, consumption pattern, and planned income growth.

MARGARET SIMERAL, Ph.D. Claremont 1975. The impact of the public employment on sex-related wage differentials.

DONALD M. SPRINGER, Ph.D. Colorado State 1975. Cost analysis systems design and computerization of early education.

JEROME STALLER, Ph.D. Temple 1975. The impact of collective bargaining on faculty at two-year public colleges.

MARY H. STEVENSON, Ph.D. Michigan 1974. Determinants of low wages for women workers.

YOSHIHIKO SUGAI, Ph.D. Iowa State 1974. A quota system policy and its impact on the labor market in the sugarcane industry; analyses through dynamic linear programming procedure, São Paulo, Brazil.

ROSA L. TIDALGO, Ph.D. Wisconsin (Madison) 1975. Wages and the wage structure in the Philippines, 1957 to 1969.

ROBERT W. TINNEY, Ph.D. Southern Methodist 1975. Changes in the industrial composition of income and the returns to skill.

SOCRATES D. TOUNTAS, Ph.D. Wayne State 1974. Internal migration and metropolitan growth: An empirical analysis.

T. JAMES TRUSSELL, Ph.D. Princeton 1975. Selected applications of model fertility schedules.

WALTER A. VERDON, Ph.D. Nebraska 1974. Rates of return to investment in education at all levels in the state of Nebraska.

CLAIR B. VICKERY, Ph.D. Maryland 1973. Why unemployment rates differ by race and sex.

DORIS K. WALKER, Ph.D. California (Davis) 1975. Factors affecting the supply of labor in selected paramedical occupations in a local labor market.

ROBIN J. C. WALTHER, Ph.D. California (Berkeley) 1975. Occupational position and urban economic structure: A study of the employment experience of foreign immigrants.

RICHARD E. WEBER, Ph.D. Rutgers 1975. The labor market for computer user personnel.

EDWARD T. WILLAUER, Ph.D. Illinois 1974. The demand for youth labor: A cross-sectional approach.

STANLEY C. WISNIEWSKI, Ph.D. Catholic 1975. Factors in the growth of unionism—1956–66.

LOUISE B. WOLITZ, Ph.D. California (Berkeley) 1975. An analysis of the labor market for policemen.

HAROLD H. WONG, Ph.D. California (Berkeley) 1975. The relative economic status of Chinese, Japanese, and black men in California.

HARRIET S. ZELLNER, Ph.D. Columbia 1975. The determinants of the occupational distribution of women.

GILROY J. ZUCKERMAN, Ph.D. North Carolina State 1974. An economic study concerned with the valuation of human life in court cases involving the accidental death of females.

Welfare Programs; Consumer Economics; Urban and Regional Economics

JACK E. ADAMS, Ph.D. Oklahoma State 1975. Low-rent public housing in Oklahoma: Case studies of benefits and costs in selected city sizes.

RONALD G. ADAMS, Ph.D. Kansas State 1975. A study of the relationship between economic and environmental variables for a region.

SAMIR FOUAD AL-KHURI, Ph.D. Vanderbilt 1975. The economics of condominium and cooperative housing in the United States.

CHARLES E. ANDERSON, Ph.D. Stanford 1975. An evaluation of Medicaid as a scheme for redistributing physician services towards the medically indigent and socially disadvantaged.

DONALD M. ATWATER, Ph.D. California (Los Angeles) 1974. Environmental housing services in a two-good metropolitan housing market framework.

MARCY C. E. AVRIN, Ph.D. Stanford 1975. Some economic effects of residential zoning in San Francisco.

CHRISTOPHER T. BABB, Ph.D. Washington (St. Louis) 1974. The household's evaluation of commuting time and the urban housing market.

BOYCE B. BALFOUR II, Ph.D. North Carolina State 1974. Spatial planning and analysis of the location of public school facilities.

DAVID M. BARTON, Ph.D. Virginia 1974. Alternative institutional arrangements for medical care insurance.

John H. Bishop, Ph.D. Michigan 1974. The private demand for places in higher education: A study of the response of Project Talent high school students to variations in the cost of attending college.

Robert J. Caswell, Ph.D. Michigan 1975. Studies in the production of health.

Jay G. Chambers, Ph.D. Stanford 1975. The impact of collective negotiations for teachers on resource allocation in public school districts.

William J. Chambers, Ph.D. Columbia 1974. The optimal allocation of land to transportation in urban areas.

Jon B. Christianson, Ph.D. Wisconsin (Madison) 1974. Physician location and the public interest.

James P. Clay, Ph.D. Kansas State 1974. Factors causing changes in the undergraduate enrollment at the Kansas Regents' institutions of higher education.

Steven A. Cobb, Ph.D. Brown 1975. A theory of local government and residential choice.

Larry R. Davis, Ph.D. Arkansas 1975. An appraisal of selected economic development projects of the Texarkana, Texas model cities program.

James J. Doyle, Ph.D. Pennsylvania State 1974. The intraregional distribution of new jobs according to migratory status, 1965–70.

Arthur L. Ekholm, Jr., Ph.D. Oklahoma State 1975. Regional economic adjustment to the depletion of groundwater and petroleum: High plains of Oklahoma and Texas.

George B. Fallis, Ph.D. Princeton 1975. The technology of production and the location of employment in urban areas.

Sherman T. Folland, Ph.D. Iowa 1974. Nonprofit hospitals and goal orientation.

Harry E. Frech III, Ph.D. California (Los Angeles) 1974. The regulation of health insurance.

Joseph Friedman, Ph.D. California (Berkeley) 1975. Housing location and the supply of local public services.

George Galster, Ph.D. Massachusetts Institute of Technology 1974. A bid-rent analysis of housing market discrimination.

Loren V. Geistfeld, Ph.D. Minnesota 1974. A technical efficiency approach to consumer decision making.

Sharif A. Ghalib, Ph.D. Pennsylvania State 1974. Expenditure patterns of aged households in the United States: An econometric analysis.

Frank A. Giarratani, Ph.D. West Virginia 1975. Essays in regional input-output analysis.

Janet C. Goulet, Ph.D. Notre Dame 1975. Credit availability to women.

Margaret E. W. Greenbaum, Ph.D. Maryland 1974. The poverty problem: A merit want approach.

Stuart J. Greenfield, Ph.D. Texas (Austin) 1975. An analysis of nonmetropolitan poverty reduction: A cross-sectional analysis of economic growth, industrialization, and labor mobility in selected MCA hinterlands, 1959–69.

Leland V Gustafson, Ph.D. Florida State 1974. The theoretical and empirical effects of city migration on the benefit-cost ratios for impure public goods.

Hatim M. Hajj, Ph.D. South Carolina 1975. The feasibility of rapid transit in medium-size urban areas.

David Harrison, Ph.D. Harvard 1974. The cost-benefit distribution of federal automobile emission standards.

Heidi I. Hartmann, Ph.D. Yale 1974. Consumer products, technology, and women's work in the home, 1900–30 in the United States.

Dennis R. Heffley, Ph.D. California (Santa Barbara) 1975. Assignment problems and efficient spatial allocation.

Paul E. Hodges, Ph.D. Stanford 1975. A theory on the consumer demand for durable goods.

Martin R. Holmer, Ph.D. Massachusetts Institute of Technology 1975. The economic and political causes of the "welfare crisis."

Alexander B. Holmes, Ph.D. State University of New York (Binghamton) 1975. An empirical estimation of a cost function for elementary and secondary education.

Richard J. Horn, Ph.D. Iowa State 1975. Population, land use, and income in a central place region.

Joseph C. Hu, Ph.D. Oklahoma State 1974. A study of intercounty commuting in Oklahoma, 1970.

Antoine G. Jabbour, Ph.D. Catholic 1975. The interrelationships among size expenditure level and quality in nonpublic schools: The case of Seventh Day Adventists.

Eugene W. Johnson, Ph.D. Virginia 1974. A model of slow adjustment to relative price differences in the housing market.

David D. Jones, Ph.D. Indiana 1975. Two projections of housing demand and supply: 1970–2020.

Frank S. Jones, Ph.D. McMaster 1974. Macro-economic influences on urban employment patterns: An input-output analysis.

Kevin C. Kelley, Ph.D. Brown 1975. Urban amenities and disamenities: A critique of and a contribution towards the Tobin-Nordhaus measure of economic welfare.

David M. Kiefer, Ph.D. Michigan 1974. The impact of neighborhood conditions on optimization in housing supply.

John E. Kushman, Ph.D. North Carolina 1974. Prices and outputs in the voluntary hospital.

Paul S. Lande, Ph.D. Southern California 1974. Regional growth in the United States: A production function approach.

Michael Lawson, Ph.D. Iowa 1975. An economic investigation of physicians' practice arrangements with stochastic processes.

Robert M. Leekley, Ph.D. Michigan State 1974. A multiple output approach to public education.

Keith R. Leitner, Ph.D. Kansas State 1974. An analysis of state programs for pollution control as a factor influencing industrial location.

JOSEPH LIPSCOMB, Ph.D. North Carolina 1975. Specifications and estimations of a production function for dental services.

IRWIN LITHWICK, Ph.D. Western Ontario 1974. The growth of urban population in Canada to 1976: An economic model.

GEORGE R. McDOWELL, Ph.D. Michigan State 1975. Whose preferences count? A study of the effects of community size and characteristics on the distribution of the benefits of schooling.

KANITTA MEESOOK, Ph.D. Massachusetts Institute of Technology 1974. Studies in the theory of wealth and income distribution.

ROBERT A. MOFFITT, Ph.D. Brown 1975. The effect of employment decentralization on job opportunity in the central city, 1960–70.

JAMES R. MOOR, JR., Ph.D. Wayne State 1975. An economic demographic forecasting model for the Detroit region.

DICKIE R. MURPHY, Ph.D. Iowa State 1974. Equality of educational opportunity and educational finance: A question of the applicability of the Serrano v. Priest decision.

PHILIP MUSGROVE, Ph.D. Massachusetts Institute of Technology 1974. Determination and distribution of permanent household income in urban South America.

JOEL L. NAROFF, Ph.D. Brown 1975. Choice of mode in an urban environment.

RONALD NEEDLEMAN, Ph.D. City (New York) 1973. Effects of modifications of political boundaries upon fiscal flows: A study of a hypothetical case.

DAVID NELSON, Ph.D. Oregon 1975. The income elasticity of the demand for housing service.

HOWARD F. NEWELL, Ph.D. Indiana 1974. The measurement of urban decentralization with special emphasis on the decentralization of manufacturing employment within the Chicago area, 1890–1930.

STANLEY NOVAL, Ph.D. Princeton 1975. The demand for international tourism and travel: Theory and measurement.

NAOHIRO OGAWA, Ph.D. Hawaii 1975. The impact of the aging of the Japanese population upon government pension schemes.

ROBERT F. O'NEIL, Ph.D. Fordham 1975. Local property tax rates and intraregional plant location: A case study in Connecticut 1958–67.

LARRY J. OZANNE, Ph.D. Stanford 1975. An analysis of urban housing markets experiencing population growth.

ORMAN H. PAANANEN, Ph.D. Brown 1975. Efficiency and welfare aspects of the urban housing insurance market: 1950–72.

JAN PALMER, Ph.D. Michigan State 1974. An economic analysis of sentencing and recidivism in the Michigan criminal justice system.

CAROL D. PFROMMER, Ph.D. Texas (Austin) 1975. Alternative growth strategies for eastern Kentucky.

BARRY M. POPKIN, Ph.D. Cornell 1974. Vitamin A deficiency in the Philippines: The development and analysis of alternate interventions.

FERDINAND J. POST, Ph.D. Oregon 1974. The conversion of recreational land into homesite development and the role of the landowner in controlling it: A case study employing theoretical/empirical economic methods.

YVON PROULX, Ph.D. Michigan State 1974. The location pattern of the largest manufacturing corporations and its impact on rural area economic growth: A case analysis for the states of Michigan and California.

JAMES R. RAMSEY, Ph.D. Kentucky 1974. The social return on selected programs at the University of Kentucky.

MICHAEL H. REDWINE, Ph.D. Oregon 1975. The determinants of residential land values: A case study of the Eugene-Springfield area of Oregon.

WILLIAM M. RHODES, Ph.D. Minnesota 1974. An economic analysis of the criminal courts.

STEVEN M. ROCK, Ph.D. Northwestern 1975. The redistributive effects of mass transit in the Chicago area.

DONALD F. SCOTT, Ph.D. Missouri 1974. Estimation of regional export labor income multipliers for Missouri.

PAUL SERVELLE, Ph.D. Claremont 1975. Effects of pricing policies on enrollment demand and the distribution of enrollments between the private and public sectors of higher education.

GERALD L. SETTER, Ph.D. Minnesota 1974. The hours of work supply decision: A study of metropolitan and rural Minnesota nurses.

BENJAMIN H. SEXAUER, JR., Ph.D. Stanford 1974. The role of habit formation and inventory adjustment in a dynamic demand model.

JEAN A. SHACKELFORD, Ph.D. Kentucky 1974. An inquiry into the growth characteristics of intermediate sized Standard Metropolitan Statistical Areas.

PAULO R. SILVA, Ph.D. Michigan State 1974. Estimating least cost human diets in the northeast of Brazil with stochastic programming—theoretical issues and policy implications.

MARGARET C. SIMMS, Ph.D. Stanford 1975. A decision model for intradistrict allocation of educational resources.

WILLIAM H. SMITH, Ph.D. Kansas State 1975. An analysis of state aid to local school districts for pupil transportation in Kansas with proposals for modification.

DOUGLAS O. STEWART, Ph.D. Michigan 1974. The value of single family residential property: An empirical investigation.

THOMAS TABASZ, Ph.D. State University of New York (Binghamton) 1975. Toward an economics of prisons.

AMY R. TAYLOR, Ph.D. Harvard 1975. The demand for labor in nonprofit institutions: Studies of the hospital industry.

MICHAEL D. TOYER, Ph.D. Duke 1975. Health care delivery and a price mechanism for resource allocation.

RONALD L. TROSPER, Ph.D. Harvard 1974. The economic impact of the allotment policy on the Flathead Indian Reservation.

JOHN VAHALY, JR., Ph.D. Vanderbilt 1975. The location of service and office activities in Nashville-Davidson County, 1970.

ALLEN VANDER MEULEN, JR., Ph.D. Brown 1975. Economic efficiency and residential segregation by household income: An evaluation of selected mechanisms for assigning neighbors to households.

ERNESTO C. VENEGAS, Ph.D. Minnesota 1974. Simulation of a regional economy—a systems approach to migration.

DAVID P. VOGT, Ph.D. Syracuse 1974. An interregional analysis of the empirical importance of factor intensity reversals.

WALTER H. VOGT, Ph.D. Syracuse 1974. Income distribution of fiscal centralization: A comparative analysis of the effects in ten *U.S.* central cities.

THOMAS J. WALSH, Ph.D. California (Berkeley) 1975. An exploration of the effects of sponsored research on higher education.

IRWIN J. WARGON, Ph.D. California (Los Angeles) 1974. An economic analysis of blood outdating in the United States.

ANDREW T. WILLIAMS, Ph.D. Stanford 1975. An economic analysis and evaluation of California's probation subsidy program: A case study of a performance grant-in-aid program.

RONALD L. WILLIAMS, Ph.D. California (Santa Barbara) 1974. Outcome-based measurements of medical care output: The case of maternal and infant health.

ROBERT H. WILSON, Ph.D. Yale 1975. The methodological defects of "urban dynamics" and their implications for the study of social systems.

JOHN M. YINGER, Ph.D. Princeton 1975. Three essays on the economics of discrimination in housing.

JAMES A. ZINGALI, Ph.D. Florida State 1975. A theoretical and empirical analysis of the relationship between city size and income.

ANNOUNCEMENT

NOTICE TO ALL GRADUATE DEPARTMENTS

The December 1976 issue of the *Review* will carry the seventy-third list of doctoral dissertations in political economy in American universities and colleges. The list will specify doctoral degrees conferred during the academic year terminating June 1976. This announcement is an invitation to send us information for the preparation of the list. This announcement supersedes and replaces a letter which was sent annually from the managing editor's office.

The *Review* will publish in its December 1976 issue the names of those who will have been awarded the doctoral degree since June 1975 and the titles of their dissertations. Dissertation abstracts will no longer be published, as these are published elsewhere.

By June 30, please send us this information on 3×5 cards, conforming to the style shown below, one card for each individual. Please indicate by a classification number in the right-hand corner the field in which the thesis should be classified. The classification system is that used by the *Journal of Economic Literature* and printed in every issue.

JEL Classification No. ________

Name: LAST NAME IN CAPS: First Name, Initial

Institution Granting Degree: ________

Degree Conferred (Ph.D. or D.B.A.) ________ Year ________

Dissertation Title: ________

When degrees in economics are awarded under different names, such as Business Administration, Public Administration, or Industrial Relations, candidates in these fields whose training has been ***primarily in economics*** should be included.

All items and information should be sent to the Assistant Editor, *American Economic Review*, Box Q, Brown University, Providence, Rhode Island 02912.

THE AMERICAN ECONOMIC REVIEW

VOLUME LXV

BOARD OF EDITORS

Bela Balassa
Eugene F. Fama
Martin S. Feldstein
Ann F. Friedlaender
Stephen M. Goldfeld
Robert J. Gordon
Robert E. Hall
Bent Hansen
James R. Melvin
William D. Nordhaus
Leonard A. Rapping
Stephen Resnick
Anna J. Schwartz
Jerome Stein
Joseph E. Stiglitz
S. C. Tsiang
Finis Welch
Marina v. N. Whitman

MANAGING EDITOR
GEORGE H. BORTS

THE AMERICAN ECONOMIC ASSOCIATION
Executive Office: Nashville, Tennessee
Editorial Office: Brown University, Providence, Rhode Island

Copyright 1975

American Economic Association

CONTENTS OF ARTICLES AND SHORTER PAPERS

W. W. Heller: What's Right With Economics? 1

R. B. Freeman: Supply and Salary Adjustments to the Changing Science Manpower Market: Physics, 1948–1973 27

B. B. Aghevli: The Balance of Payments and the Money Supply Under the Gold Standard Regime: U.S. 1879–1914 40

R. M. Coen: Investment Behavior, The Measurement of Depreciation, and Tax Policy 59

M. S. Feldstein: Wealth Neutrality and Local Choice in Public Education 75

W. L. Gramm: Household Utility Maximization and the Working Wife 90

O. H. Schenone: A Dynamic Analysis of Taxation 101

D. A. Graham: A Geometrical Exposition of Input-Output Analysis 115

L. De Alessi: Toward an Analysis of Postdisaster Cooperation 127

J. M. Buchanan and G. Tullock: Polluters' Profits and Political Response: Direct Controls Versus Taxes 139

T. T. Su: Durability of Consumption Goods Reconsidered 148

J. G. Haworth, J. Gwartney, and C. Haworth: Earnings, Productivity, and Changes in Employment Discrimination During the 1960's 158

G. de Menil and S. S. Bhalla: Direct Measurement of Popular Price Expectations 169

G. A. Calvo: Efficient and Optimal Utilization of Capital Services 181

R. E. Baldwin: International Trade and Economic Growth: A Diagrammatic Analysis 187

R. Hendler: Lancaster's New Approach to Consumer Demand and Its Limitations 194

E. James: A Note on Uncertainty and the Evaluation of Public Investment Decisions 200

R. Eisner and P. J. Lawler: Tax Policy and Investment: An Analysis of Survey Responses 206

N. S. Barrett: The U.S. Phillips Curve and International Unemployment Rate Differentials: Comment 213

R. J. Flanagan: Reply 222

L.-S. Fan: On the Reward System 226

D. Orr: The Industrial Composition of U.S. Exports and Subsidiary Sales to the Canadian Market: Comment 230

T. Horst: Reply 235

C. R. Link: Black Education, Earnings, and Interregional Migration: A Comment and Some New Evidence 236

L. Weiss and J. G. Williamson: Even Newer Evidence 241

R. M. Korkie: Theory of the Firm Facing Uncertain Demand: Comment 245

H. E. Leland: Reply 248

M. C. Kemp and E. Tower: Nontraded Goods, Factor Market Distortions, and the Gains from Trade: Comment 249

R. N. Batra: Reply 251

E. S. Mills and F. M. Peterson: *Environmental Quality*: The First Five Years 259

E. F. Fama: Short-Term Interest Rates as Predictors of Inflation 269

J. E. Stiglitz: The Theory of "Screening," Education, and the Distribution of Income 283

P. J. Lloyd: Substitution Effects and Biases in Nontrue Price Indices 301

M. K. Block and J. M. Heineke: A Labor Theoretic Analysis of the Criminal Choice 314

R. A. Meyer: Monopoly Pricing and Capacity Choice Under Uncertainty 326

M. N. Baily: Dynamic Monopsony and Structural Change 338

D. A. Wise: Academic Achievement and Job Performance 350

L. R. Christensen, D. W. Jorgenson, and L. J. Lau: Transcendental Logarithmic Utility Functions 367

R. B. Westin: Empirical Implications of Infrequent Purchase Behavior in a Stock Adjustment Model 384

I. Ehrlich: The Deterrent Effect of Capital Punishment: A Question of Life and Death 397

L. G. Reynolds: China as a Less Developed Economy 418

P. S. Heller: A Model of Public Fiscal Behavior in Developing Countries: Aid, Investment, and Taxation 429

E.-M. Claassen: Demand for International Reserves and the Optimum Mix and Speed of Adjustment 446

M. Allais: The Hereditary and Relativistic Formulation of the Demand for Money: Circular Reasoning or a Real Structural Relation? 454

J. L. Scadding: Professor Allais' Theory of the the Demand for Money: Rejoinder 465

I. Friend and R. Westerfield: Required Disclosure and the Stock Market: Comment 467

G. J. Benston: Rejoinder 473

M. E. Canes: The Simple Economics of Incentive Contracting: Note 478

W. Hamovitch and R. D. Morgenstern: The Principal Cause of Academic Salary Differentials: Research Output or Experience?: Comment 484
D. A. Katz: Reply 486
J. M. Barron: Segmentation of the Labor Market: Comment 487
C. C. Holt: Rejoinder 491
G. Hanoch: The Elasticity of Scale and the Shape of Average Costs 492
W. L. Holahan: The Welfare Effects of Spatial Price Discrimination 498
W. A. Bomberger: Wage Determination, Inflation, and the Industrial Structure: Comment 504
S. A. Ross and M. L. Wachter: Reply 507
G. Smith: Pitfalls in Financial Model Building: A Clarification 510
D. R. Bohi and G. W. Scully: Buyer's Prices, Seller's Prices, and Price Flexibility: Comment 517
G. J. Stigler: Reply 526
H. O. Carter: The 1975 Report of the President's Council of Economic Advisers: Food and Agriculture 533
W. Poole: The 1975 Report of the President's Council of Economic Advisers: Long on Analysis, Short on Policy 539
J. Niehans: Interest and Credit in General Equilibrium with Transactions Costs 548
K. Lancaster: Socially Optimal Product Differentiation 567
T. W. Epps: Security Price Changes and Transaction Volumes: Theory and Evidence 586
M. Paglin: The Measurement and Trend of Inequality: A Basic Revision 598
P. A. Meyer and J. A. Neri: A Keynes-Friedman Money Demand Function 610
I. Friend and C. Lieberman: Short-Run Asset Effects on Household Saving and Consumption: The Cross-Section Evidence 624
A. G. Schweinberger: Pure Traded Intermediate Products and The Heckscher-Ohlin Theorem 634
W. L. Springer: Did The 1968 Surcharge Really Work? 644
J. A. Ordover and E. S. Phelps: Linear Taxation of Wealth and Wages for Intragenerational Lifetime Justice: Some Steady-State Cases 660
J. A. Frenkel and C. A. Rodriguez: Portfolio Equilibrium and the Balance of Payments: A Monetary Approach 674
R. E. Wong: Profit Maximization and Alternative Theories: A Dynamic Reconciliation .. 689
M. J. Beckmann: The Limits to Growth in a Neoclassical World 695
M. B. McElroy and J. C. Poindexter: Capital Gains and the Aggregate Consumption Function: Comment 700
K. B. Bhatia: Reply 704
E. J. Mishan: The Concept and Measure of Consumer's Surplus: Comment 708
M. E. Burns: Reply 710
G. E. Mumy and S. H. Hanke: Public Investment Criteria for Underpriced Public Products 712
P. R. P. Coelho: Externalities, Liability, Separability, and Resource Allocation: Comment 721
A. Gifford, Jr. and C. C. Stone: Comment ... 724
P. Greenwood, C. Ingene, and J. Horsfield: Comment 728
J. R. Marchand and K. P. Russell: Reply 730
P. Bohm: Option Demand and Consumer's Surplus: Comment 733
R. Schmalensee: Reply 737
A. Daniere: The Optimum Lifetime Distribution of Consumption Expenditures: Comment 740
L. C. Thurow: Reply 753
E. G. Davis and J. E. Olson: A Model of a Highway System Financed by a Trust Fund 756
D. Weber: Optimal Advertising: A Note 759
L. H. Meyer and W. R. Hart: On the Effects of Fiscal and Monetary Policy: Completing the Taxonomy 762
P. L. Cheng: Default Risk, Scale, and the Homemade Leverage Theorem: Note 768
J. R. Crotty and L. A. Rapping: The 1975 Report of the President's Council of Economic Advisers: A Radical Critique 791
M. Rubinstein: Securities Market Efficiency in an Arrow-Debreu Economy 812
F. Modigliani and E. Tarantelli: The Consumption Function in a Developing Economy and The Italian Experience 825
W. E. Weber: Interest Rates, Inflation, and Consumer Expenditures 843
R. Dornbusch: Exchange Rates and Fiscal Policy in a Popular Model of International Trade 859
B. J. Moore: Equities, Capital Gains, and the Role of Finance in Accumulation 872
D. F. Bradford: Constraints on Government Investment Opportunities and the Choice of Discount Rate 887
I. Friend and M. E. Blume: The Demand for Risky Assets 900

A. Razin and U. Ben-Zion: An Intergenerational Model of Population Growth 923

J. M. Barron: Search in the Labor Market and the Duration of Unemployment: Some Empirical Evidence 934

R. Deacon and P. Shapiro: Private Preference for Collective Goods Revealed Through Voting on Referenda 943

C. R. Hulten: Technical Change and the Reproducibility of Capital 956

G. R. Faulhaber: Cross-Subsidization: Pricing in Public Enterprises 966

H. Hori: Revealed Preference for Public Goods 978

O. Hochman: Market Equilibrium Versus Optimum in a Model with Congestion: Note 992

G. Daly and J. F. Giertz: Externalities, Extortion, and Efficiency 997

W. W. Brown and M. O. Reynolds: A Model of IQ, Occupation, and Earnings 1002

R. Carson: On Monopoly Welfare Losses: Comment 1008

D. A. Worcester, Jr.: Comment 1015

A. Bergson: Reply 1024

D. A. Graham and J. M. Vernon: The Economics of the Network-Affiliate Relationship: Comment 1032

S. M. Besen and R. Soligo: Reply 1037

◆

CONTENTS OF PAPERS AND PROCEEDINGS

Richard T. Ely Lecture

A. M. Rivlin: Income Distribution—Can Economists Help? 1

The Economics of Labor: An Assessment of Recent Research

G. G. Cain: The Challenge of Dual and Radical Theories of the Labor Market to Orthodox Theory 16

G. E. Johnson: Economic Analysis of Trade Unionism 23

H. G. Lewis: Economics of Time and Labor Supply 29

Discussion by **P. E. Stephan** 35

Past Developments and Future Prospects for Ethnic Minority Groups

R. W. Fogel: Three Phases of Cliometric Research on Slavery and its Aftermath 37

B. E. Anderson and P. A. Wallace: Public Policy and Black Economic Progress: A Review of the Evidence 47

D. H. Swinton: Factors Affecting the Future Economic Prospects of Minorities 53

Discussion by **R. L. Trosper and D. Fusfeld** .. 59

The Human Capital Approach: An Appraisal

F. Welch: Human Capital Theory: Education, Discrimination, and Life Cycles 63

S. Bowles and H. Gintis: The Problem with Human Capital Theory—A Marxian Critique 74

Supply and Mobility of Women Economists

A. H. Amsden and C. Moser: Job Search and Affirmative Action 83

M. H. Strober: Women Economists: Career Aspirations, Education, and Training 92

B. B. Reagan: Two Supply Curves for Economists? Implications of Mobility and Career Attachment of Women 100

Joint Luncheon Address

H. J. Witteveen: Inflation and the International Monetary Situation 108

Understanding World Inflation

W. H. Branson: Monetarist and Keynesian Models of the Transmission of Inflation 115

G. L. Perry: Understanding World Inflation .. 120

R. T. Seldon: Monetary Growth and the Long-Run Rate of Inflation 125

R. L. Teigen: Interpreting Recent World Inflation 129

Domestic Adjustment to Balance of Payments Disequilibrium

M. v.N. Whitman: The Payments Adjustment Process and the Exchange Rate Regime: What Have We Learned? 133

Discussion by **R. Dornbusch, W. J. Fellner, A. Kafka, and P. A. Volcker** 147

25 Years After the Rediscovery of Money: What Have We Learned?

S. Fischer: Recent Developments in Monetary Theory 157

W. C. Brainard and R. N. Cooper: Empirical Monetary Macroeconomics: What Have We Learned in the Last 25 Years? 167

Discussion by **M. Friedman and F. Modigliani** 176

A Critical Look at the Keynesian Model: Theory and Application

R. Clower and A. Leijonhufvud: The Coordination of Economic Activities: A Keynesian Perspective 182

R. Eisner: The Keynesian Revolution Reconsidered 189

J. Tobin: Keynesian Models of Recession and Depression 195

Personal Savings in the Postwar World: Implications for the Theory of Household Behavior

F. T. Juster and L. D. Taylor: Towards a Theory of Saving Behavior 203

B. Strumpel: Saving Behavior in Western Germany and the United States 210

M. R. Darby: Postwar U.S. Consumption, Consumer Expenditures, and Saving 217

Discussion by **T. Mayer** 223

Microeconomic Theory: Conflict and Contract

J. M. Buchanan: A Contractarian Paradigm for Applying Economic Theory 225

D. K. Foley: Problems vs. Conflicts: Economic Theory and Ideology 231

A. K. Klevorick: Law and Economic Theory: An Economist's View 237

Sources and Uses of Panels of Microdata

H. S. Parnes: The National Longitudinal Surveys: New Vistas for Labor Market Research 244

J. N. Morgan: Using Survey Data from the University of Michigan's Survey Research Center 250

S. Kallek: Potential Applications of Census Bureau Economic Series in Microdata Analysis 257

K. Dickinson and H. W. Watts: The Experimental Panel Data Resources at the Poverty Institute Data Center.................... 263

The Behavior of Firms in Oligopolistic Markets: In Theory and Practice

P. L. Joskow: Firm Decision-Making Processes and Oligopoly Theory.............. 270

M. Shubik: Oligopoly Theory, Communication, and Information.................. 280

Emerging Structural Changes in Transport and Public Utilities

A. L. Morton: Northeast Railroads: Restructured or Nationalized?................... 284

G. S. Goldstein and L. N. Moses: Transportation Controls and the Spatial Structure of Urban Areas.......................... 289

P. L. Joskow and P. W. MacAvoy: Regulation and the Financial Condition of the Electric Power Companies in the 1970's........... 295

A Reassessment of Development Economics

I. Adelman: Development Economics—A Reassessment of Goals.................... 302

H. B. Chenery: The Structuralist Approach to Development Policy.................... 310

S. A. Resnick: State of Development Economics.............................. 317

What Have We Learned from Economic Growth Theory?

R. Findlay: Implications of Growth Theory for Trade and Development................ 323

D. J. Harris: The Theory of Economic Growth: A Critique and Reformulation.... 329

R. R. Nelson and S. G. Winter: Growth Theory from an Evolutionary Perspective: The Differential Productivity Puzzle...... 338

China and India: Development During the Last 25 Years

B. Richman: Chinese and Indian Development: An Interdisciplinary Environmental Analysis.............................. 345

T. E. Weisskopf: China and India: Contrasting Experiences in Economic Development. 356

Discussion by **P. Desai and J. G. Gurley**..... 365

East-West Technological Transfer Issues and Problems

E. Mansfield: International Technology Transfer: Forms, Resource Requirements, and Policies.............................. 372

E. A. Hewett: The Economics of East European Technology Imports from the West... 377

T. G. Rawski: Problems of Technology Absorption in Chinese Industry............. 383

Taxation of Natural Resources

M. Gillis and C. E. McLure, Jr.: Incidence of World Taxes on Natural Resources with Special Reference to Bauxite............ 389

G. M. Brannon: U.S. Taxes on Energy Resources.............................. 397

Discussion by **A. W. Wright**................ 405

Optimizing Models of Public Decision Making

J. A. Ferejohn and M. P. Fiorina: Purposive Models of Legislative Behavior........... 407

M. J. Roberts: An Evolutionary and Institutional View of Public and Private Companies 415

The Principles Course: What Should Be in It and Where Should It Be Going?

K. E. Boulding: Some Observations on the Learning of Economics................. 428

J. G. Gurley: Some Comments on the Principles Course.......................... 431

W. L. Hansen: New Approaches to Teaching the Principles Course.................... 434

CONTRIBUTORS TO ARTICLES AND SHORTER PAPERS

Aghevli, B. B. 40
Allais, M. 454
Baily, M. N. 338
Baldwin, R. E. 187
Barrett, N. S. 213
Barron, J. M. 487, 934
Batra, R. N. 251
Beckmann, M. J. 695
Benston, G. J. 473
Ben-Zion, U. 923
Bergson, A. 1024
Besen, S. M. 1037
Bhalla, S. S. 169
Bhatia, K. B. 704
Block, M. K. 314
Blume, M. E. 900
Bohi, D. R. 517
Bohm, P. 733
Bomberger, W. A. 504
Bradford, D. F. 887
Brown, W. W. 1002
Buchanan, J. M. 139
Burns, M. E. 710
Calvo, G. A. 181
Canes, M. E. 478
Carson, R. 1008
Carter, H. O. 533
Cheng, P. L. 768
Christensen, L. R. 367
Claassen, E.-M. 446
Coelho, P. R. P. 721
Coen, R. M. 59
Crotty, J. R. 791
Daly, G. 997
Daniere, A. 740
Davis, E. G. 756
Deacon, R. 943
De Alessi, L. 127
de Menil, G. 169
Dornbusch, R. 859
Ehrlich, I. 397
Eisner, R. 206
Epps, T. W. 586
Fama, E. F. 269
Fan, L.-S. 226
Faulhaber, G. R. 966
Feldstein, M. S. 75
Flanagan, R. J. 222
Freeman, R. B. 27
Frenkel, J. A. 674
Friend, I. 467, 624, 900
Giertz, J. F. 997
Gifford, A., Jr. 724
Graham, D. A. 115, 1032
Gramm, W. L. 90
Greenwood, P. 728
Gwartney, J. 158
Hamovitch, W. 484
Hanke, S. H. 712
Hanoch, G. 492
Hart, W. R. 762
Haworth, C. 158
Haworth, J. G. 158
Heineke, J. M. 314
Heller, P. S. 429
Heller, W. W. 1
Hendler, R. 194
Hochman, O. 992
Holahan, W. L. 498
Holt, C. C. 491
Hori, H. 978
Horsfield, J. 728
Horst, T. 235
Hulten, C. R. 956
Ingene, C. 728
James, E. 200
Jorgenson, D. W. 367
Katz, D. A. 486
Kemp, M. C. 249
Korkie, R. M. 245
Lancaster, K. 567
Lau, L. J. 367
Lawler, P. J. 206
Leland, H. E. 248
Lieberman, C. 624
Link, C. R. 236
Lloyd, P. J. 301
McElroy, M. B. 700
Marchand, J. R. 730
Meyer, L. H. 762
Meyer, P. A. 610
Meyer, R. A. 326
Mills, E. S. 259
Mishan, E. J. 708
Modigliani, F. 825
Moore, B. J. 872
Morgenstern, R. D. 484
Mumy, G. E. 712
Neri, J. A. 610
Niehans, J. 548
Olson, J. E. 756
Ordover, J. A. 660
Orr, D. 230
Paglin, M. 598
Peterson, F. M. 259
Phelps, E. S. 660
Poindexter, J. C. 700
Poole, W. 539
Rapping, L. A. 791
Razin, A. 923
Reynolds, L. G. 418
Reynolds, M. O. 1002
Rodriguez, C. A. 674
Ross, S. A. 507

Rubinstein, M. 812
Russell, K. P. 730
Scadding, J. L. 465
Schenone, O. H. 101
Schmalensee, R. 737
Schweinberger, A. G. 634
Scully, G. W. 517
Shapiro, P. 943
Smith, G. 510
Soligo, R. 1037
Springer, W. L. 644
Stigler, G. J. 526
Stiglitz, J. E. 283
Stone, C. C. 724
Su, T. T. 148
Tarantelli, E. 825
Thurow, L. C. 753
Tower, E. 249
Tullock, G. 139
Vernon, J. M. 1032
Wachter, M. L. 507
Weber, D. 759
Weber, W. E. 843
Weiss, L. 241
Westerfield, R. 467
Westin, R. B. 384
Williamson, J. G. 241
Wise, D. A. 350
Wong, R. E. 689
Worcester, D. A., Jr. 1015

CONTRIBUTORS TO PAPERS AND PROCEEDINGS

Adelman, I. 302
Amsden, A. H. 83
Anderson, B. E. 47
Boulding, K. E. 428
Bowles, S. 74
Brainard, W. C. 167
Brannon, G. M. 397
Branson, W. H. 115
Buchanan, J. M. 225
Cain, G. G. 16
Chenery, H. B. 310
Clower, R. 182
Cooper, R. N. 167
Darby, M. R. 217
Desai, P. 365
Dickinson, K. 263
Dornbusch, R. 147
Eisner, R. 189
Fellner, W. J. 148
Ferejohn, J. A. 407
Findlay, R. 323
Fiorina, M. P. 407
Fischer, S. 157
Fogel, R. W. 37
Foley, D. K. 231
Friedman, M. 176
Fusfeld, D. 60
Gillis, M. 389
Gintis, H. 74
Goldstein, G. S. 289
Gurley, J. G. 368, 431
Hansen, W. L. 434
Harris, D. J. 329
Hewett, E. A. 377
Johnson, G. E. 23
Joskow, P. L. 270, 295
Juster, F. T. 203
Kafka, A. 151
Kallek, S. 257
Klevorick, A. K. 237
Leijonhufvud, A. 182
Lewis, H. G. 29
MacAvoy, P. W. 295
McLure, C. E., Jr. 389
Mansfield, E. 372
Mayer, T. 223
Modigliani, F. 179
Morgan, J. N. 250
Morton, A. L. 284
Moser, C. 83
Moses, L. N. 289
Nelson, R. R. 338
Parnes, H. S. 244
Perry, G. L. 120
Rawski, T. G. 383
Reagan, B. B. 100
Resnick, S. A. 317
Richman, B. 345
Rivlin, A. M. 1
Roberts, M. J. 415
Selden, R. T. 125
Shubik, M. 280
Stephan, P. E. 35
Strober, M. H. 92
Strumpel, B. 210
Swinton, D. H. 53
Taylor, L. D. 203
Teigen, R. L. 129
Tobin, J. 195
Trosper, R. L. 59
Volcker, P. A. 153
Wallace, P. A. 47
Watts, H. W. 263
Weisskopf, T. E. 356
Welch, F. 63
Whitman, M. v.N. 133
Winter, S. G. 338
Witteveen, H. J. 108
Wright, A. W. 405

Join the American Economic Association

Membership includes:

- a subscription to both *The American Economic Review* (quarterly) plus May, *Papers and Proceedings* and the *Journal of Economic Literature* (quarterly)
- annual ballots and nomination forms for elected officers of the Association
- preregistration material and announcement of the annual meeting of the Allied Social Science Associations
- announcements of special interest to members, such as the life insurance program started in 1973; new book offerings, etc.

1976 ANNUAL MEMBERSHIP RATES

Class 1	Regular member with rank of assistant professor or lower or annual income of $12,000 or less	$25.00
Class 2	Regular member with rank of associate professor or annual income of $12,000-$20,000	$30.00
Class 3	Regular member with rank of full professor or annual income above $20,000	$35.00
Class 4	Junior member (registered students; certification must be submitted yearly) Student status must be certified by your major professor or school registrar; please use the space provided below	$12.50
	Members in classes listed above living in a foreign country should include one of the following amounts for postage; $2.10 P.U.A.S. countries or $2.50/all other countries and Canada.	
Class 5	Family member (second membership without publications; two or more living at same address)	$ 5.00
	Dues payable in U. S. currency only.	

(Under the terms of a resolution passed at the annual meeting of members of the American Economic Association on Dec. 29, 1973, $20.00 of the annual membership dues is considered as payment for the AER and JEL; the remainder is considered a contribution to exempt functions of the Association.)

Name in full ______________________

Address ______________________

____
City State/Country Zip/Post Code

Endorsed by (AEA member) ______________________

For junior members only: I certify that ______________________

is enrolled as a student at ______________________

____
Authorized signature

Please send with remittance to: American Economic Association
1313 21st Avenue South
Nashville, Tennessee 37212

What Samuelson knows about economics...could fill ten editions, and does!

Professionals do not reach the zenith of their field by resting on their laurels. Neither does a textbook. That's why Paul A. Samuelson's ECONOMICS, now in its Tenth Edition, has continued to grow, reflecting the author's continued growth.

Concentrating on the last quarter of the twentieth century, its problems and their implications, the Tenth Edition is filled with new insights on energy. Professor Samuelson has personally scrutinized the entire book, particularly the material on OPEC countries; ecology; shortages; inflation and recession; and the Federal Reserve. And the major statistical tables, charts, and textual facts were expertly updated by Peter Temin, Economic Historian and professor at MIT.

Samuelson is a respected book because it *has* changed with the times. For example:

- The emphasis of the microeconomic chapter on agriculture has been shifted from farm price supports to the new world-wide scare of population pressure, possible adverse climatic change, and apprehensions that the terms of trade may shift in coming decades in favor of Iowa and prime farm regions and against industrial and developing nations.
- The revised chapters on macroeconomics now consider the fact that the world-wide recession of 1973-75, the worst in the post World War II age of Keynes, has been primarily characterized by "stagflation", or stagnation with inflation. The monetarist debate is also summarized.
- Part Six's policy formulation grapples with the new facts concerning energy scarcity and OPEC cartel oil pricing.
- This edition concentrates on problems concerning the quality of life, sex and racial discrimination, and comparative economic systems in respect to their ideology on equality and incentives. The worsening Phillips curve and the rate of deteriorating national unemployment in the world's affluent countries are also treated.

Prices subject to change without notice.

ECONOMICS TENTH EDITION

Paul A. Samuelson,
Massachusetts Institute of Technology
1976, 960 pages, $12.95 (tent.)

Accompanied by—

Instructor's Manual

Test Bank, free to adopters

STUDY GUIDE TO ACCOMPANY SAMUELSON: ECONOMICS,
Tenth Edition
Romney Robinson, 1976, 327 pages (tent.),
$5.50

ECONOMICS: AN INTRODUCTORY PROGRAM FOR USE WITH SAMUELSON,
Tenth Edition
Edward Foster, 1973, 380 pages, $4.95 soft

READINGS IN ECONOMICS,
Seventh Edition
Edited by **Paul A. Samuelson,** 1973, 384 pages,
$7.95c; soft cover, $5.95

College Division
McGRAW-HILL BOOK COMPANY
1221 Avenue of the Americas
New York, New York 10020

MACROECONOMICS: THE THEORY OF INCOME, EMPLOYMENT, AND THE PRICE LEVEL

Michael R. Darby, University of California, Los Angeles, 1976, 480 pages (tent.), $12.95

Reflecting developments in macroeconomic theory over the past fifteen years, this text presents a new, teachable synthesis of modern macroeconomics within a dynamic framework. Its central theme is that macroeconomic shocks begin processes of adjustment toward long-run growth equilibriums, processes which may take several years.

MACROECONOMICS, Fifth Edition

Thomas F. Denburg, American University, and **Duncan M. McDougall,** University of Kansas and Queens University, 1976, 528 pages (tent.), $12.50 (tent.)

Completely revised and updated, the Fifth Edition of this popular intermediate macroeconomics text takes a contemporary look at inflationary-recession and problems of policy mix. It also comprehensively discusses international policy problems and the policy choices available to economies under new international arrangements.

BETTER ECONOMICS BOOKS

SOLUTIONS: A NEW ECONOMIC HISTORY OF AMERICA

Gerarld Gunderson, North Carolina State University, 1976, 640 pages (tent.), $11.95 (tent.)

This comprehensive, up-to-date text excellently applies economic analysis to help readers understand economic behavior—past and present. The broad approach includes the entire span of American history, starting with the economies of the pre-Columbian Indians and ending with modern economic developments after World War II.

AN INTRODUCTION TO MODERN THEORIES OF ECONOMIC GROWTH

Hywel G. Jones, Oxford University, 1976, 304 pages (tent.), $14.95 (tent.)

Systematic and precise, this intermediate text surveys the principal theories and models of economic growth. Emphasizing economic rather than mathematical problems, the text shows how these theories are linked with other areas of economics and discusses why strong controversies have occurred.

MONEY AND BANKING

Dudley G. Luckett, Iowa State University, 1976, 544 pages (tent.), $13.95 (tent.)

Covering material previously available only in professional journals, this text uses regression analysis to treat economic theory and describe economic institutions. Factual material is arranged in a cause-and-effect framework, permitting economists to test theoretical hypotheses against factual evidence.

COLLEGE DIVISION **McGRAW-HILL BOOK COMPANY** 1221 Avenue of the Americas, New York, New York 10020

PUBLIC FINANCE IN THEORY AND PRACTICE, Second Editon

Richard A. Musgrave, Harvard University, and **Peggy B. Musgrave,** Northeastern University, 1976, 768 pages (tent.), $14.50

The Second Edition of this widely acclaimed text combines sound economic analysis with rich empirical and institutional material. Discussing such major topcis as the proper size of the public sector, issues of income redistribution, welfare and social security, among others, the text examines the economics of fiscal policies in conjunction with politcal and social forces surrounding them.

THE ECONOMICS OF POVERTY

Richard Perlman, University of Wisconsin, 1976, 224 pages (tent.), $4.95 (tent.) soft

Presenting a balanced, objective view of the economics of poverty in the United States, this text closely relates poverty to economic institutions within the framework of background, causes, and corrective measures. It uses elementary data analysis to explain and exemplify the theoretical and conceptual issues examined.

ECONOMETRIC MODELS AND ECONOMIC FORECASTS

Robert S. Pindyck, Massachusetts Institute of Technology, and **Daniel L. Rubinfeld,** University of Michigan, 1976, 704 pages (tent.), $14.95 (tent.)

Introducing the science and art of building models, the authors help readers determine the type of model to build, test the model statistically, and apply it to practical forecasting and analytic problems.

MAKE BETTER ECONOMISTS!

THE NEW INDUSTRIAL ORDER: CONCENTRATION, REGULATION, AND PUBLIC POLICY

Samuel R. Reid, University of New Hampshire, 1976, 256 pages (tent.), $8.95 (tent.)

Focusing on microeconomic developments within industrial and financial organization, Professor Reid examines economic concentration as a comprehensive phenomenon, treating both international and national dimensions. Concerned with issues, he examines a variety of empirical studies, and makes public policy suggestions covering federal, state, and local levels of action.

REGULATION: A CASE APPROACH

Leonard W. Weiss, University of Wisconsin, Madison, and **Allyn Strickland,** University of Wisconsin, 1976, 224 pages (tent.), $5.95, soft cover

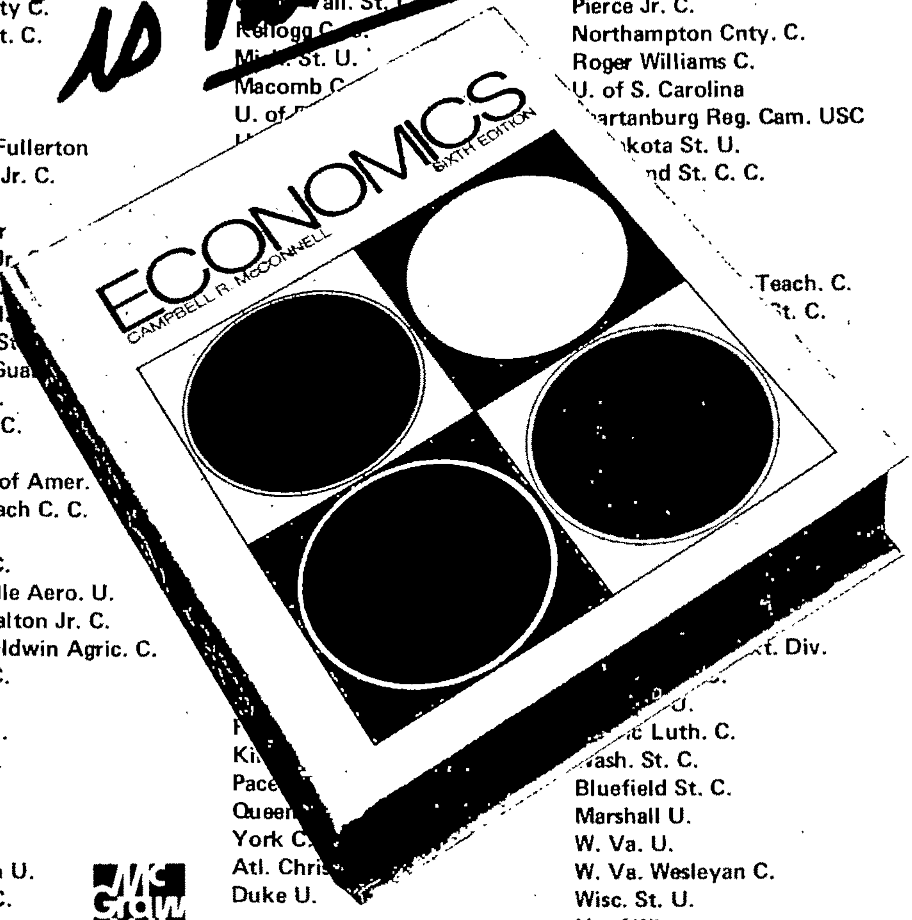
This is the first junior/senior level case book in regulation. Broad in scope, its case studies include the constitutional and economic basis for control, regulatory rate-making (and marginal cost pricing), regulation in transportation, and others. All case studies are carefully excerpted and prefaced by an introduction for each major set of cases.

Troy St. C.
Jefferson St. Jr. C.
Calhoun St. C. C.
U. of Ala.
A.T. Brewer St. Jr. C.
Chattahoochee Vall. C.
Ariz. St. U.
Mesa C. C.
U. o k
He de
trus
East L
El Ca C.
Fuller n Jr. C
L.A. City C.
Cal. St. C., L.A.
Napa C.
Orange Coast C
Riverside City C.
San Diego St. C.
Cal. St. U.
Salano C. C.
Ventura C.
Cal. St. C., Fullerton
Saddlebuck Jr. C.
Regis C.
U. of Denver
Araphahoe Jr.
W. Conn. St
U. of New H
Cen. Conn. St
U.S. Coast Gua
Tunxia C. C.
Hartford C. C.
U. of Del.
Catholic U. of Amer.
Daytona Beach C. C.
U. of Fla.
Brevard C. C.
Embry Riddle Aero. U.
Okaloosa Walton Jr. C.
Abraham Baldwin Agric. C.
Albany Jr. C.
Augusta C.
Columbus U.
U. of Idaho
Chi. St. U.
Elgin C. C.
Ill. St. U.
Ill. Wesleyan U.
Morton Jr. C.

W. Ill. U.
Harper C.
Linc. Land C. C.
Moraine V C.
Thornton
Indian Hi C.
Grac an C
C.
n. U.
Loyola U.
U. New O eans
U. Me and
Ca nsv
t. C.
oke C. C.
Nichols C.
Qui si
f ech.
all. St.
Kellogg C
St. U.
Macomb C
U. of
F
Ki
Pace
Queen
York C
Atl. Chri
Duke U.

High Point C.
Gaston C.
Cleveland St. C.
Kent St. U.
Ohio St. U.
Ohio U.
Sinclair C.
U. of Akron
Franklin U.
Cen. Oklahoma St. U.
Conn s St. Agric. C.
tte C.
LaS e C.
Pen St. U.
V nova U.
Bucks Cnty
ny Cnty. C. C.
Westmoreland City C. C.
Pierce Jr. C.
Northampton Cnty. C.
Roger Williams C.
U. of S. Carolina
artanburg Reg. Cam. USC
kota St. U.
nd St. C. C.
Teach. C.
t. C.
t. Div.
U.
c Luth. C.
ash. St. C.
Bluefield St. C.
Marshall U.
W. Va. U.
W. Va. Wesleyan C.
Wisc. St. U.
U. of Wisc.

McConnell's Number 6 is number 1



College Division
McGRAW-HILL BOOK COMPANY 1221 Avenue of the Americas, New York, New York 10020

New Textbooks from
ACADEMIC PRESS

**ECONOMICS:
PRIVATE AND PUBLIC CHOICE**

JAMES D. GWARTNEY
Florida State University

An important new principles textbook that focuses on giving the student an understanding of the economic way of thinking. Professor Gwartney applies economic analysis to modern economic issues and supports these applications with timely examples, real world supportive data, and presentations of both the strengths and weaknesses of policy alternatives.

In addition to the traditional market analysis, Gwartney thoroughly treats collective decision-making, and compares and contrasts the two processes.

With supplementary materials:
COURSEBOOK—A student guide with self-tests and readings.
TEST BANK—A large selection of class-tested multiple choice questions.

Ready February, 1976   about 700 pages

Available now:

**MONEY AND INFLATION:
A MONETARIST APPROACH**

J. HUSTON McCULLOCH
Boston College

A new supplementary textbook examining the theoretical and practical aspects of the monetary theory of inflation from an unabashedly monetarist point of view.

121 pages, $3.95

For complimentary copies of the above two textbooks, write to the Textbook Department. Please indicate course enrollment and present textbook.

ACADEMIC PRESS
A Subsidiary of Harcourt Brace Jovanovich, Publishers
111 Fifth Avenue, New York, N.Y. 10003

Important New Texts

ARCHIBALD and LIPSEY
AN INTRODUCTION TO MATHEMATICAL ECONOMICS:
METHODS AND APPLICATIONS
March 1976. Tentative: 480 pages.

BRANSON and LITVACK
MACROECONOMICS
February 1976. Tentative: 416 pages. Instructor's Manual.

BRESSLER
A UNIFIED INTRODUCTION TO MATHEMATICAL ECONOMICS
August 1975. Tentative: 667 pages. Solutions Manual.

GITMAN
PRINCIPLES OF MANAGERIAL FINANCE
February 1976. Tentative: 672 pages. Study Guide. Instructor's Manual.

GITMAN, JOHNSON, and FLAHERTY
CASES IN MANAGERIAL FINANCE
April 1976. Tentative: 256 pages. Instructor's Manual.

KAISH
MICROECONOMICS:
LOGIC, TOOLS, AND ANALYSIS
February 1976. Tentative: 448 pages. Instructor's Manual.

LEVITAN, MANGUM, and MARSHALL
HUMAN RESOURCES AND LABOR MARKETS:
LABOR AND MANPOWER IN THE AMERICAN ECONOMY
SECOND EDITION
January 1976. Tentative: 672 pages.

To request examination copies, write to Lilian Schein, Dept. 147, Harper & Row, 10 East 53d Street, New York, N.Y. 10022. Please include course title, enrollment, and present text.

LIPSEY and STEINER
ECONOMICS
FOURTH EDITION
January 1975. 964 pages. Instructor's Manual. Study Guide and Problems. Test Bank.

MORGAN
ECONOMIC DEVELOPMENT:
CONCEPT AND STRATEGY
July 1975. 429 pages.

NORTH and MILLER
THE ECONOMICS OF PUBLIC ISSUES
THIRD EDITION
January 1976. Tentative: 179 pages; paper. Instructor's Manual.

THORN
INTRODUCTION TO MONEY AND BANKING
January 1976. Tentative: 544 pages. Instructor's Manual.

YEAGER
INTERNATIONAL MONETARY RELATIONS:
THEORY, HISTORY, AND POLICY, SECOND EDITION
January 1976. Tentative: 670 pages.

YOTOPOULOS and NUGENT
ECONOMICS OF DEVELOPMENT:
EMPIRICAL INVESTIGATIONS
February 1976. Tentative: 416 pages.

1817 Harper & Row

The ECONOMICS INSTITUTE

SPECIAL PROPAEDEUTIC TRAINING AND ORIENTATION PROGRAM FOR

INTERNATIONAL Graduate Students in ECONOMICS and RELATED FIELDS

UNIVERSITY OF COLORADO, JANUARY 19 - AUGUST 18, 1976

With six consecutive sessions beginning January 19, February 23, March 29, May 3, June 7 and July 12, 1976.

The Institute is specially designed for students from other countries who are planning to undertake graduate studies at universities in the United States in economics and related fields, including business and agricultural economics. An intensive and specialized *English* instructional program is progressively substituted by an expanding component of preparatory course offerings in *mathematics, statistics* and *economic theory,* and lectures and research seminars on special field topics including agriculture, business and development. The program also includes field trips and social recreational activities designed to acquaint the participants with campus and community life in the United States.

Admissions to the sessions beginning in January, February, March, and May, are restricted to students needing further intensive training in English; those who would score in excess of 500 on the TOEFL at the point of registration should apply for the later (summer) sessions.

Universities and sponsoring agencies are invited to utilize the Institute's Credentials Evaluation Service and its Spring and Summer testing and training programs in their evaluation of candidates for graduate program admissions and related financial support.

Further information and application materials may be obtained from the DIRECTOR, ECONOMICS INSTITUTE, UNIVERSITY OF COLORADO, BOULDER, COLORADO 80302, U.S.A.

Program sponsored by the

American Economic Association

in cooperation with

The American Assembly of Collegiate Schools of Business

Good business from Wiley/Hamilton.

MACROECONOMICS:

Measurement, Theory, and Policy
By Michael C. Lovell,
Wesleyan University

This is an outstanding text for intermediate and upper-level courses in macroeconomics. Lovell stresses modern theoretical advances in analyzing unemployment, inflation, balance of payments, disequilibrium, business cycles, growth.

Schematic flowcharts show students how the equations which constitute the building blocks of a sequence of models fit together. Each chapter ends with a list of key concepts, exercises, and problems.

1975 $14.95

ECONOMICS AND SOCIETY

By Leonard W. Weiss,
University of Wisconsin, Madison

Some economics books stand out. Among these few, Weiss is simply outstanding.

ECONOMICS AND SOCIETY builds its analysis around supply and demand, and aggregate supply and demand. But the theory remains as straightforward as possible—never presented for its own sake. Rather, concepts are applied to such policy issues as monopoly, discrimination, labor market, balance of payments, and public goods. While kept elementary, every concept is developed in the context of realistic, practical experience...to give your students the major elements of economics in a meaningful, realistic setting.

1975 $10.95

MACROECONOMICS:

Money, Prices, and Income
By Dudley W. Johnson,
University of Washington

This book presents an integrated approach to macroeconomics, that is, it covers the traditional subject matter, as well as the economics of money and banking, and international monetary economics. It also provides an exhaustive analysis of monetarism versus fiscalism.

November 1976 $11.95 (tent.)

Forthcoming:

MICROECONOMICS: Theory and Applications

By Robert Y. Awh,
Mississippi State University

Awh offers a balanced, comprehensive coverage of up-to-date intermediate microeconomics, combining a solid treatment of theory with real-world applications. 1976

MACROECONOMICS, Third Edition

By John Lindauer,
Murray State University, Kentucky

This is a thorough, straightforward, and graphic treatment of macroeconomics. This edition has been revised to provide an up-to-date analysis of the state of the art.

February 1976

INTRODUCTION TO ECONOMICS: an issues approach

By Chris Rhoden, Solano College

Pedagogically sound, extremely readable: Rhoden's INTRODUCTION TO ECONOMICS features objectives, summaries, glossaries, and discussion questions for each chapter—as well as related readings drawn from Business Week, Wall Street Journal, etc. Also included: brain teasers (projects for student research).

This introduction to basic concepts and contemporary issues is perfect for the one-term basic economics course. Available with an Instructor's Manual and a Student Study Guide.

February 1976 $9.95 (tent.)

For your complimentary examination copies, contact your Wiley representative or write to George Thomsen, Dept. 033, at the address below. Please include course title, enrollment and present text.

WILEY/HAMILTON
The Business and Economics
Publishing Group of Wiley,
1129 State Street,
Santa Barbara, California 93101

Prices subject to change without notice. A5353-HP

Please mention THE AMERICAN ECONOMIC REVIEW *When Writing to Advertisers*

For Your Students —

For your Money and Banking students the KEY is ...

Money, Banking and Economic Analysis

Thomas D. Simpson
Macalester College

— a distinctive blend of theory and application which balances institutional and historical coverage with economic analysis. Stresses the explanatory nature of economic principles. Emphasizes relevant current topics including: inflation, inflationary recession, international monetary reform, budget deficits, and many more. Includes chapter summaries, review questions, and ample references for further study.

1975 496 pp. cloth $12.95

For examination copies of these important new texts, return this coupon to Robert Jordan, Box CJ-272, College Division, Prentice-Hall, Englewood Cliffs, N.J., 07632.

☐ **Money, Banking and Economic Analysis** by Thomas D. Simpson
☐ **Microeconomics for Business Decisions** by J.M. Heineke

NAME/TITLE ____________
COLLEGE ____________
STREET/BUILDING ____________
CITY ____________
STATE ____________ ZIP ________

Prices Subject to Change Without Notice

For Possible Adoption

COURSE TITLE: ____________
DATE COURSE WILL BE OFFERED: ____________

For Personal Reference

☐ Bill me
☐ Payment herewith

Keys To Opportunity

For your
Microeconomics
and Applied
Economics
students
the KEY is ...

Microeconomics for Business Decisions: Theory and Applications

J.M. Heineke
University of Santa Clara

— a striking integration of theory and application emphasizing the use of microeconomics in real-world decision situations. Explores thoroughly problems confronting managers in private sector. Integrates classical production theory and modern process-focused production theory in treating the design of the optimal production plan. Includes numerous examples and problem sets in each chapter. Appendix material reviews all math necessary to use the text.
January 1976 288 pp. cloth $13.95

From
Prentice-Hall

Pergamon Press announces PD644A

A New International Journal

Accounting, Organizations and Society

Editor-in-chief: **Dr. Anthony G. Hopwood,** *Administrative Staff College, Henley-on-Thames, England*
supported by a strong international editorial board

ACCOUNTING, ORGANIZATIONS AND SOCIETY is an international journal devoted to the behavioral, organizational and social aspects of accounting. Contributions will range from the original theoretical and empirical to review articles describing the state of the art in specific areas. Shorter critical assessments of experiments related to the behavioral and social aspects of accounting will also be published, as well as book reviews. Published quarterly Subscription rate $60.00 2 years 1976/77 $114.00

Free specimen copy available on request

READINGS IN MANAGERIAL ECONOMICS

Edited by **I.B. Ibrahim, K.K. Seo and P.G. Vlachos,** *Department of Business Economics and Quantitative Methods, University of Hawaii*

Combines articles selected from over fifty leading professional publications. The selection has been designed to show the application of theoretical concepts to business decision-making. The articles, grouped into five main subject sections, are self-contained but combine to form a very useful managerial economic text with a careful balance between theory and application. 435 pp approx 76 illus

ISBN 0 08 019605 5 $19.50 approx ISBN 0 08 019604 7 flexicover $14.50 approx Spring 1976

TECHNOLOGICAL CHANGE: ECONOMICS, MANAGEMENT AND ENVIRONMENT Edited by **Bela Gold,** *Professor of Economics, Case Western Reserve University*

A contributory volume from Case Western Reserve University on aspects of industrial change and the interaction between technological change and economic and social relationships.

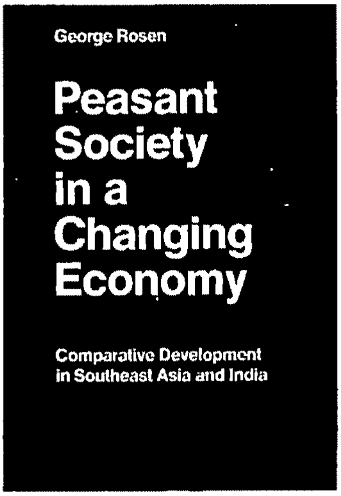
188 pp 31 illus September 1975 ISBN 0 08 018012 4 $15.00

Pergamon Press
Maxwell House, Fairview Park, Elmsford, New York 10523
Headington Hill Hall, Oxford OX3 0BW, England

PEASANT SOCIETY IN A CHANGING ECONOMY

Comparative Development in Southeast Asia and India

GEORGE ROSEN



George Rosen

Peasant Society in a Changing Economy

Comparative Development in Southeast Asia and India

Stressing the possibility of revolution in the Philippines, Indonesia, and Thailand — unless significant change in their governments occurs — George Rosen suggests ways to avoid widespread human misery and social and political instability in Southeast Asia. In this interdisciplinary study, he relates the process of social and political change to that of economic development and compares the experiences of these three countries with those of India. ". . . embodies wide-ranging and imaginative scholarship on a series of very significant problems." — Milton J. Esman, Center for International Studies at Cornell University. 269 pages. **$8.95**

UNIVERSITY OF ILLINOIS PRESS URBANA 61801

Important New Titles
From Lexington Books

- AIR POLLUTION CONTROL
  A Welfare Economic Interpretation
  Robert Kohn, *Southern Illinois University* 192 pp. 1975 $16.50
- DISCRETIONARY CONTROL OF AGGREGATE ECONOMIC ACTIVITY
  Kenneth D. Garbade, *New York University* 208 pp. 1975 $17.50
- ECONOMIC ASPECTS OF THE ENERGY CRISIS
  Harry W. Richardson, *University of Southern California* 256 pp. 1975 $16.00
- POPULATION AND DEVELOPMENT IN SOUTHEAST ASIA
  edited by **John F. Kantner** and **Lee McCaffrey**,
  The Johns Hopkins University 352 pp. 1975 $15.00
- PUBLIC UTILITY REGULATION
  Change and Scope
  edited by **Werner Sichel**, *Western Michigan University;* and
  Thomas G. Gies, *University of Michigan* 128 pp. 1975 $12.50

The Economics of International Relations & Trade

- LIQUIDITY CREATION AND DISTRIBUTION IN THE EUROCURRENCY MARKETS
  John Hewson 192 pp. 1975 $16.00
- THE OIL SECURITY SYSTEM
  An Important Strategy for Achieving Oil Security and Reducing Oil Prices
  D.H. Newlon, *National Science Foundation* and
  N.V. Breckner, *Federal Energy Administration* 128 pp. 1975 $9.50
- TRADE INSTABILITY AND ECONOMIC DEVELOPMENT
  An Empirical Study
  Odin Knudsen and **Andrew Parnes**, *San Jose State College* 160 pp. 1975 $14.00
- U.S. TRADE IN THE SIXTIES AND SEVENTIES
  Continuity and Change
  edited by **Kenneth Jameson** and **Roger Skurski**, *University of Notre Dame*
  160 pp. 1974 $12.50

The Economics of Health

- THE COST OF HEALTH INSURANCE ADMINISTRATION
  An Economic Analysis
  Roger D. Blair, *University of Florida;* and
  Ronald J. Vogel, *Social Security Administration* 208 pp. 1975 $16.50
- HEALTH ECONOMICS IN DEVELOPING NATIONS
  Alan Sorkin, *Johns Hopkins University* In Press 1976

Lexington Books
125 Spring Street
Lexington, MA 02173

New Text Emphasizes Applications

Liberal use of applications intermixed with theory

PRICES AND MARKETS: MICRO-ECONOMICS

Questions and Exercises on contemporary problems and situations conclude every chapter

Students learn theory . . . and how to apply it to current problems

Charles W. Baird
California State University, Hayward
230 pages Cloth, $11.95

Write for your complimentary examination copy. Please include course, number, average enrollment, and present text.

west's college department
MARKETING
50 W. Kellogg Blvd., St. Paul, Minn. 55102

NEW FROM CALIFORNIA

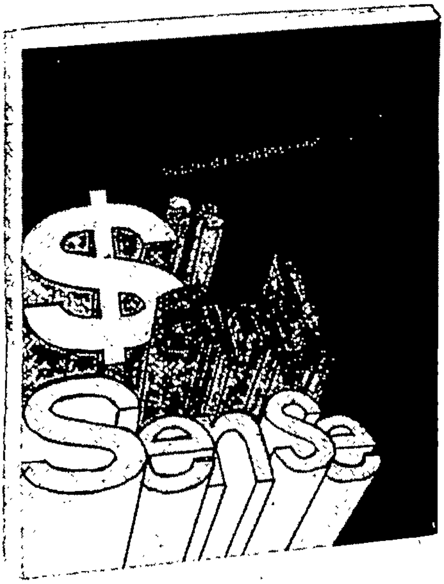
Urban Development
The Process and the Problems
WALLACE F. SMITH

This is a comprehensive textbook on urban economics that explains how the important decisions that determine the physical development of cities are made. It provides an overview of a complex, interlocking system that involves landowners, developers, investors, planning agencies, and other arms of local government. It is intended for courses in urban planning, real estate, and urban land economics.

399 pages, $14.75

At your bookstore

UNIVERSITY OF CALIFORNIA PRESS
BERKELEY 94720



Dollars and Sense
An Introduction to Economics

Marilu McCarty,
Georgia Institute of Technology

Dollars and Sense is an interesting introduction to economic reasoning and analysis presenting both micro and macro concepts through current issues and applications to everyday life. Concepts and analytical tools are presented simply, followed by examples and applications relating policies to the individual decision-making level. Each chapter includes objectives, self-tests, topics for discussion, and definitions. Special essays on economic issues appear throughout the book. Instructor's Manual with test items. January 1976, 320 pages, illustrated, paper, approx. $6.95

National Income Analysis and Forecasting

Edward J. Chambers, University of Alberta
R. Haney Scott, University of Washington
Roger S. Smith, University of Alberta
Complete Instructor's Manual
1975, 406 pages, illustrated, cloth $11.95

Problems in National Income Analysis and Forecasting

Revised Edition
R. Haney Scott, University of Washington
1972, 138 pages, illustrated, paper $4.95

Macroeconomics

Analysis and Policy
James L. Cochrane, University
of South Carolina
Samuel Gubins, Haverford College
B. F. Kiker, University of South Carolina
1974, 398 pages, illustrated, cloth $11.95

For further information write to
Jennifer Toms, Department SA
Scott, Foresman College Division
1900 East Lake Avenue
Glenview, Illinois 60025

Scott Foresman College Division

Benjamin J. Taylor and Fred Witney's LABOR RELATIONS

Revised to provide an even more effective basic text or supplement for courses in labor law, industrial relations, and government and labor. Suitable for the undergraduate liberal arts student, and the business or law student.

Flexible in format. Chronicles the dynamic history of labor relations, with focus on the environment shaping the legal structure of labor relations. To maintain student interest, the authors analyze the environment of labor relations legislation by tracing the etiology of the main issues — from their origins down through the present day.

About half the book examines the development of collective bargaining and labor relations through the Wagner Era, and then strikes a balance between historical perspectives of the law and its current problems. The student will be able to identify and trace major trends in the law of collective bargaining, the reason for these trends, and their consequences on the overall functioning of industrial relations.

Stressing the institutional framework in which collective bargaining is cast, the authors probe for the root problems inherent in government efforts to define the rights, duties, and obligations of unions and employers.

With focus on the ever changing and developing laws of labor relations — the **NEW 2ND EDITION of LABOR RELATIONS LAW** integrates into its material:

- Doctrines established by the National Labor Relations Board and the courts, with particular attention to recent NLRB applications of the Taft-Hartley Act and recent changes in the Board's jurisdiction.
- Minority group problems under Taft-Hartley and civil rights legislation.
- The "make whole" doctrine originally established in Ex-Cell-O.

LAW, 2ND EDITION, 1975

- Industrial conflict innovations.
- Developments in the public sector.
- Developments in the control of internal union affairs established by the Landum-Griffin Act.
- The *Collyer* doctrine.

RETAINED FROM THE PREVIOUS SUCCESSFUL EDITION:

- Economic analyses of problems resulting from government influence.
- Analyzes and discusses in an understandable way the major NLRB and court cases, with clear delineation of the cases' significance to employers, unions, employees, and the public.
- Offers an extensive bibliography and a full text of relevant statutes included in the appendices.
- Uses a readable style and careful organization.

1975, 745 pp., cloth (51963-7) $16.50

THE AUTHORS: BENJAMIN J. TAYLOR, University of Oklahoma and
FRED WITNEY, Indiana University.

To order your copy of Taylor and Witney's Labor Relations Law, 2nd Edition, 1975 from PRENTICE-HALL, please fill out and mail the attached coupon ...

Mail this coupon to Robert Jordan,
Prentice-Hall, Inc., Dept. J,292, Englewood
Cliffs, N.J. 07632

☐ **(51963-7) LABOR RELATIONS LAW, 2ND EDITION, 1975**
by Benjamin J. Taylor and Fred Witney
$16.50

NAME/TITLE ____________
DEPARTMENT ____________
COLLEGE ____________
CITY ____________
STATE ____________
ZIP ____________

FOR POSSIBLE ADOPTION:

COURSE TITLE ____________
DATE COURSE WILL BE OFFERED: ____________
USUAL ENROLLMENT: ____________
AUTHOR OF PRESENT TEXT: ____________

FOR PERSONAL REFERENCE:

☐ Bill me
☐ Payment herewith
Prices subject to change without notice.

gsh 9/20/75

FROM INDUSTRIAL RELATIONS SECTION, PRINCETON UNIVERSITY

PUBLIC ASSISTANCE FOR MOTHERS IN AN URBAN LABOR MARKET
by Daniel H. Saks, Michigan State University.
Testing the proposition that the welfare system has become a subsector of the economy to which potential recipients respond rationally, Professor Saks examines the trade-off between benefit levels, welfare tax rates, and labor-market opportunities. He recommends a middle ground between case-by-case discretion and a uniform negative income tax in which programs are tailored to the needs of each segment of the welfare population.
1975. 103 pp. $3.50.

PUBLIC ASSISTANCE AND WORK EFFORT: THE LABOR SUPPLY OF LOW-INCOME FEMALE HEADS OF HOUSEHOLD
by Robert G. Williams, Mathematica.
This study provides strong evidence of a causal relationship between elements of the AFDC program and the work effort of its recipients. It presents estimates of the effect on budgetary costs and labor supply of altering implicit tax rates on earnings, rates that are shown to have been considerably lower than those specified in the statutes. 1975. 128 pp. $3.50.

FORTHCOMING SPRING 1976

EVALUATING THE LABOR MARKET EFFECTS OF SOCIAL PROGRAMS. Contributions by Orley Ashenfelter, James Heckman, George Johnson, Dale Mortenson, Finis Welch, Morris Goldstein and Robert Smith. Estimated 250 pp. Approximately $5.00.

Industrial Relations Section, Princeton University
P.O. Box 248, Princeton, N.J. 08540

THE CITIZEN AND THE STATE
Essays on Regulation
George J. Stigler
These essays represent the development of Stigler's thought on the theory, history, and present practice of public regulation of economic life.
1975 224 pages Cloth $10.95

ENERGY: THE POLICY ISSUES
Edited and with a Foreword by Gary D. Eppen
With a Preface by Harold S. Geneen
Eminent scholars from a number of fields discuss the major themes of American energy policy. Published under the auspices of *The Journal of Business.*
1975 136 pages Cloth $9.50

THE UNIVERSITY
OF CHICAGO PRESS
Chicago 60637

GOVERNMENT, LABOR, AND INFLATION
Wage Stabilization in the United States
Daniel Quinn Mills
This timely work analyzes how government, labor unions, and management in the United States have dealt with the problem of controlling inflation.
1976 Cloth $16.50 January

ESSAYS ON MONEY AND INFLATION
David E. W. Laidler
The distinguished British economist David E. W. Laidler provides a new macroeconomic, monetarist approach to inflation that views the problem in terms of the economy as a whole.
1975 256 pages Cloth $15.00

Now in Paper:
Milton Friedman's Monetary Framework, A Debate With His Critics, ***Edited by Robert J. Gordon****, 192 pages/Paper $3.50.*

FROM ACADEMIC PRESS . . .

PROGRESS AGAINST POVERTY

A Review of the 1964-1974 Decade

by ROBERT D. PLOTNICK and
FELICITY SKIDMORE

A Volume in the
INSTITUTE FOR RESEARCH ON POVERTY
POVERTY POLICY ANALYSIS Series

In 1964, the war on poverty was declared. Now, more than a decade later, it is appropriate to reflect on what has been accomplished in the war against poverty during the period from 1964 to 1974. PROGRESS AGAINST POVERTY provides a comprehensive, detailed, analytical report of changes in the United States poverty population in the 1964-1974 decade. Using both absolute and relative definitions of poverty, it studies in detail the major trends in the size and demographic characteristics of the poverty population, assesses the impact of macroeconomic conditions on the extent of poverty, and analyzes the contribution of the cash transfer system toward reducing poverty.

1975, 254 pp. cloth, $12.50/£6.50;
paper, $4.50/£2.35

THE PATTERNS OF INTERRACIAL POLITICS

Conflict and Cooperation in the City

by PETER K. EISINGER

A Volume in the
INSTITUTE FOR RESEARCH ON POVERTY
MONOGRAPH Series

This book is one of the few to attempt to make explicit the implications for elite behavior of the configurations of attitudes and behaviors of ordinary citizens. Based on survey data collected in Milwaukee in 1970, it details the patterns of political behavior and belief among blacks and whites in Milwaukee, and shows how these patterns help structure the context of costs and opportunities that bear on elites as they forge relations of conflict or cooperation between majority and minority groups. It deals with elites and masses, political integration and goals, and patterns of conventional political participation, and also presents survey data on issues, such as protest and community control, not normally covered in conventional surveys.

1975, about 220 pp., $11.00/£5.70

ENTRY INTO THE AMERICAN LABOR FORCE

by MICHAEL D. ORNSTEIN

A Volume in the
QUANTITATIVE STUDIES IN SOCIAL
RELATIONS Series

ENTRY INTO THE AMERICAN LABOR FORCE analyzes the critical first years of men's work experiences. It focuses on the effects of race, family background, and education on the levels of first jobs, and, secondly, on the educational, occupational, and wage mobility which takes place during their first years in the labor force. The book is based on retrospective life histories from a national sample survey of about 1600 men between the ages of 30 and 39. In order to facilitate comparisons between whites and blacks, blacks were oversampled so they constitute almost one-half of the respondents. These interviews, which recorded the characteristics of every job held from the age of fourteen, make possible a highly detailed longitudinal analysis of job mobility.

1975, in preparation, abt. $12.00/£6.25.

SCHOOLING AND ACHIEVEMENT IN AMERICAN SOCIETY

edited by WILLIAM H. SEWELL, ROBERT M.
HAUSER, and DAVID L. FEATHERMAN

A Volume in the
STUDIES IN POPULATION Series

This outstanding collection of fifteen papers focuses on a subject of timely and pressing interest—the complex effects of schooling on socioeconomic achievement in American society. The papers offered here were originally presented at an interdisciplinary seminar on achievement processes sponsored by the American College Testing Research Institute. The book presents a wealth of recent substantive findings, and, though its primary focus is on substantive issues, nearly every chapter contains extensive methodological discussions embodying a wealth of information on the application of modern multivariate statistical techniques, with several chapters modifying on or extending existing statistical methods.

1976, in preparation, abt. $22.50/£11.70

N.B.: Postage plus 50¢ handling charge on all orders not accompanied by payment. Orders totaling less than $5.00 must be prepaid.

Prices subject to change without notice.

ACADEMIC PRESS, INC.

A Subsidiary of Harcourt Brace Jovanovich, Publishers
111 FIFTH AVENUE, NEW YORK, N.Y. 10003
24-28 OVAL ROAD, LONDON NW1 7DX

UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

Overseas opportunities in a new science and engineering University for faculty in Economics.

PhD. degree and teaching experience required. Challenging assignments in undergraduate programs. Minimum regular appointments of two years. Competitive salaries, free airconditioned furnished housing and free air transportation to and from Dhahran each two-year tour. Attractive educational assistance grants for school age children. Local transportation allowance in cash each month. All earned income without Saudi taxes. Ten month duty each year with two months vacation paid. Apply with resume including ages and sexes of children, nationality, list of references, home and office telephone numbers to:

UNIVERSITY OF PETROLEUM & MINERALS
C/O SAUDI ARABIAN EDUCATIONAL MISSION
880 THIRD AVENUE, 17th FLOOR
NEW YORK, NEW YORK 10022

Prospective candidates to be interviewed by University Recruiter early 1976.

THE URBAN INSTITUTE

NEW BOOKS

Cost Benefit Analysis & Water Pollution Policy, edited by Henry M. Peskin and Eugene P. Seskin. 19 leading experts in environmental economics present their original research and conclusions, to be used by EPA as basis for guidelines to planning agencies and states evaluating water resource programs and proposals. These studies constitute the first comprehensive survey of cost-benefit analysis techniques, capabilities and potential for antipollution programs, and provide an essential text for environmental economists and consultants. 370 pgs. $5.95 paperback; $12 clothbound.

Time of Transition: The Growth of Families Headed by Women, by Heather L. Ross and Isabel V. Sawhill. "Successfully integrates demographic, economic and sociological data on both macro- and micro-analytic levels to illuminate one of the most salient trends in our society today"—Dr. Jesse Bernard. "Potential dynamite"—Prof. Barbara Bergmann. "Valuable empirical evidence on matters most everyone has thoughts and feelings about"—Dr. Amitai Etzioni. $4.95 paperback, $10 clothbound.

2100 M Street, N.W. Washington, D.C. 20037

You don't need another journal if you already know what Policy Analysis said in its first issue about . . .

Analysis and Issues
- Scaling Welfare Benefits to Income: An Idea That Is Being Overworked, *Robert J. Lampman*
- The Correctional Carrot: Better Jobs for Parolees, *Philip J. Cook*
- Some Emerging Problems of Prepaid Health Plans in the Medi-Cal System, *Victor P. Goldberg*
- Public Policy Issues Raised by a Medical Breakthrough, *Amitai Etzioni*
- Towards a Responsible Energy Policy, *John V. Krutilla and R. Talbot Page*

Clients and Analysts
- Private Advisors and Government Policymaking, *Murray L. Weidenbaum*
- Bureaucratic Policy Analysts, *Arnold J. Meltsner*

Methods and Practices
- Applying Principles of Policy Analysis to an Offshore Oil Leasing Decision, *William R. Ahern*
- Analysis of Earnings and Income Foregone by Manpower Trainees, *Kamran Moayed-Dadkhah and Ernst W. Stromsdorfer*
- Defining Benefits of Public Programs: Some Guidance for Policy Analysts, *Robert H. Haveman and Burton A. Weisbrod*

Teaching
- A Course in the Economics of Public Policy Analysis, *Larry Seidman*

Policy Idea
- How to Save Alligators, *Steve H. Hanke*

. . . and plans to say in future issues:

The Federal Subsidy of Conventional Mass Transit, *Walter Oi*
Housing Subsidies: Do We Know What We Are Doing?, *Joseph DeSalvo*
National Health Insurance and the Sixth Stool Guaiac, *Duncan Neuhauser and A. Lewicki*
A Policy to Protect the U.S. against Oil Embargoes, *Robert Kuenne et al.*
A Critical Look at Quantitative Methodology, *Ralph Strauch*
Summary Evaluation of Delphi, *Harold Sackman*
How to Save Money with Policy Analysis, *Lewis Friedman*
Cutting the Cost of Medical Assistance, *Peter Karpoff*
The Committee on Interest and Dividends: An Assessment, *Paul Horvitz*
Policy Analysis and the Congress: An Economist's View, *Robert Haveman*
The Congressional Approach to Energy Policy, *Andrew Carron*
Food Stamps and Welfare Reform, *Richard Nathan*
Effects of Cash and In-kind Welfare Payments on Family Expenditures, *Teh-wei Hu and Norman Knaub*
Operating Subsidies for Buses, *Roger Schmenner*
Analytic Methods for Public Policy Students, *Richard Zeckhauser and Albert Nichols*

Keep on top of this rapidly changing field. Subscribe now to:

Policy Analysis

Arnold J. Meltsner, Editor

Published quarterly for the Graduate School of Public Policy, Berkeley, by the University of California Press. Annual subscription (four issues): Institutions $15, Individuals $12, Students $8. Send order with advance payment to: Policy Analysis, University of California Press, Berkeley, CA 94720.

To: **POLICY ANALYSIS**
University of California Press
Berkeley, California 94720

Start my subscription to Policy Analysis NOW.
My subscription classification is:

☐ Institution ($15) ☐ Individual ($12)
☐ Student ($8)

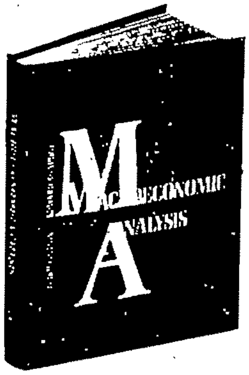
Name: ____________________

Address: ____________________

City: __________ State: ________ Zip: ________

Please find payment enclosed.

A balanced treatment of theory, measurement, and policy



MACROECONOMIC ANALYSIS

THIRD EDITION

EDWARD SHAPIRO, *University of Toledo*

The third edition of this outstanding textbook continues to provide the intermediate-level student with balanced coverage of the three major aspects of macroeconomics: theory, measurement, and policy. Professor Shapiro covers all the fundamental topics of macroeconomics, developing them at sufficient length to make them comprehensible to the average student and relying on carefully written prose and clear, explicit diagrams as his main tools of explanation. This Third Edition incorporates important new material and reflects recent developments in the field of macroeconomics. It includes a new examination of the shortcomings of the Gross National Product as a measure of national welfare; an expanded treatment of the Phillips curve in the chapter on "The Theory of Inflation"; expansion and reorganization of the material on monetary policy and fiscal policy into separate chapters; an analysis and comparison of the monetarist and Keynesian schools of thought; and a new discussion of the relation between economic growth and environmental quality.

502 pages

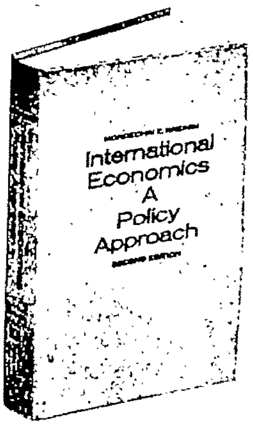
Accompanied by

STUDENT WORKBOOK

EDWARD SHAPIRO, *University of Toledo*
STEVEN J. WEISS, *University of Toledo*
BRUCE GENSEMER, *Kenyon College*

Paperbound. 300 pages

A new edition reflecting new economic realities



INTERNATIONAL ECONOMICS

A Policy Approach

SECOND EDITION

MORDECHAI E. KREININ,
Michigan State University

The new second edition of this highly respected textbook continues to provide a comprehensive, truly international introduction to world finance and trade—one that emphasizes policy and the analysis of general principles throughout. Professor Kreinin thoroughly covers present currency arrangements, analyzes the balance-of-payments problem, offers suggestions for reforms of the monetary system, explains the principles governing world trade, discusses the role of trade organizations, and analyzes selected trade problems among developing nations. This edition has been thoroughly revised throughout to reflect recent developments in international economics: new to this edition are a chapter on international financial developments in the 1970s (including the current oil crisis); a chapter on the trade problems of developing nations (including the 1974 GSP of the European Communities and the proposed GSP of the United States); and sections on the new international currency system and Eurodollars, trade relations of the expanded European community, taxation and the multinational corporation, and the monetary approach to devaluation.

473 pages

HBJ **Harcourt Brace Jovanovich, Inc.** *New York · Chicago · San Francisco · Atlanta*

EQUITY IN HEALTH SERVICES
Empirical Analyses in Social Policy

Ronald Andersen, Center for Health Administration Studies, University of Chicago, Joanna Kravits, Massachusetts Hospital Association, and Odin W. Anderson, Center for Health Administration Studies, University of Chicago, Editors

Economic and social issues determining the utilization and distribution of health care services.
288 pp $15.00

ECONOMIC PLANNING, EAST AND WEST

Morris Bornstein, University of Michigan

An authoritative analysis and evaluation of national economic planning in Eastern centrally-administered economies and Western market-oriented economies
416 pp $16.00

INPUT-OUTPUT ANALYSIS IN PUBLIC EDUCATION

Elchanan Cohn, University of South Carolina; and Stephen D. Millman, Maryland State Board for Community Colleges

Techniques for assisting educational managers in the use of input-output analysis for education.
192 pp $14.50

PETROLEUM DISCOVERIES AND GOVERNMENT POLICY
An Econometric Study of Supply

Dennis Epple, Carnegie-Mellon University

Analysis and proposals for changes in government policies affecting production and prices of petroleum to influence new discoveries of crude oil and natural gas. 128 pp $15.00

STUDIES IN NONLINEAR ESTIMATION

Stephen M. Goldfeld and Richard E. Quandt, Princeton University, Editors

New applications of techniques for econometric estimation — switching regressions; production functions; complete systems of demand functions; and models incorporating limited dependent variables.
144 pp $15.00

WHO PAYS FOR CLEAN AIR:
The Cost and Benefit Distribution of Federal Automobile Emission Standards

David Harrison, Jr., Harvard University

Quantitative estimates of the costs and benefits of current federal automobile emission control strategy for various income groups and geographical areas.
192 pp $14.00

For information on these and related Ballinger titles, please write for our catalog.

AGRICULTURAL TRACTORS:
A World Industry Study

Robert T. Kudrle, Center of International Studies, University of Minnesota

A comparative analysis of the structure, conduct and performance of the tractor industry in seven major non-Communist nations. 288 pp $16.50

NON-NUCLEAR FUTURES:
The Case for an Ethical Energy Strategy

Amory B. Lovins and John Price (co-published with Friends of the Earth International)

A closely focused outline of the main technical, ethical, social, economic and policy issues regarding the use of nuclear power.
144 pp (cloth) $11.95 (paper) $5.95

TRANSPORTATION ECONOMICS

Herbert Mohring, University of Minnesota

Suggestions for the application of basic economic theories of price and value to transportation problems. 128 pp $12.50

ADVANCES IN INPUT-OUTPUT ANALYSIS

Karen Polenske, Urban Studies & Planning, Massachusetts Institute of Technology; and Jiri Skolka, Austrian Institute for Economic Research, Vienna, Editors

Proceedings of the Sixth International Conference on Input-Output Techniques, Vienna, 1974.
464 pp $25.00

THE COSTS OF CONGESTION:
An Econometric Analysis of Wilderness Recreation

V. Kerry Smith, State University of New York at Binghamton; and Charles J. Cicchetti University of Wisconsin

Problems of management and resource allocation in low density recreational lands — general methodology for modeling the costs of congestion.
112 pp $12.50

THE DEMAND FOR CHILDREN:
The Economics of Fertility in the United States

Boone A. Turchi, University of North Carolina, Chapel Hill

The first attempt to formulate and statistically test a unified socioeconomic model of fertility in the United States. 256 pp $15.00

HOW TO SAVE GASOLINE:
Public Policy Alternatives for the Automobile

Sorrell Wildhorn, et al., The Rand Corporation

Analytical tools for evaluating public policies for conserving energy used by automobiles — compares various proposals for fuel economy. (A Rand Corporation Research Study) 228 pp $15.00

Ballinger
PUBLISHING COMPANY
17 Dunster Street, Harvard Square, Cambridge, MA 02138

TOUCHSTONE

PAPERBACKS

SIMON AND SCHUSTER
630 FIFTH AVENUE
NEW YORK, N. Y. 10020

For complimentary examination copies, write on your departmental letterhead to School and Library Services, address above. Include details on course in which you are considering using title requested, and advise us where you have seen announcement.

Hayek, F. A.
STUDIES IN PHILOSOPHY, POLITICS AND ECONOMICS

Selections from Hayek's work of the last twenty years include essays on rationalism, the philosophy of Hume, the moral element in free enterprise, planning and inflation, unions and profits, the corporation in a democratic society, and a reassessment of the author's monumental work *The Road to Serfdom.* "Dr. Hayek's scholarship is impeccable." —George Lichtheim, *New Statesman.*

☐ 20246 Index $2.95

Kindleberger, Charles P.
ECONOMIC GROWTH IN FRANCE AND BRITAIN: 1851–1950

"A synthesis of the many scattered empirical studies of economic development in the Western world examined in the light of current models . . . a most knowledgeable . . . most readable book." —P. Sargant Florence, *Annals of the American Academy of Political and Social Sciences.*

☐ 20293 Index $2.95

Green, Timothy
THE WORLD OF GOLD: The History, the Lure, and the Power of Man's Most Precious Metal

In this unique history of fact and fantasy are the fascinating details of the Great Gold Rushes of the nineteenth century, the world of private hoarders and gold smugglers, the intrigue of the gold markets of London, Zurich, Hong Kong, and Macao, the use of gold in the industrial arts. Journalist Green presents the mania, the science, the legends of man's insatiable lust for gold and explains the qualities, the curious chemistry, the power and influence that have made this yellow metal run red with blood.

☐ 20352 Index $2.45

Myrdal, Gunnar
THE POLITICAL ELEMENT IN THE DEVELOPMENT OF ECONOMIC THEORY

Translated by Paul Streeten

[Myrdal's] "avowed object is to show that most influential economists who wrote before the time of Max Weber believed that value judgments could be deduced from a study of the 'facts,' and hence that 'we must look upon the majority of modern economic doctrines as modified reminiscences of very old political thinking.' " —*The Journal of Political Economy.*

☐ 20390 Index $2.45

ADVANCED TEXTBOOKS IN ECONOMICS

Editors: C. J. BLISS, *UK and* M. D. INTRILIGATOR, *USA*
Advisory Editors: S. M. GOLDFELD, *USA;* L. JOHANSEN, *Norway;* D. W. JORGENSON, *USA;* M. C. KEMP, *Australia;* J. C. MILLERON, *France*

The aim of the series is to cover topics in economics, mathematical economics and econometrics, at a level suitable for graduate students or final year undergraduates specializing in economics.

Vol. 1: VARIATIONAL METHODS IN ECONOMICS

by G. HADLEY *and* M. C. KEMP, *School of Economics, The University of New South Wales, Australia*
1971, 388 pages. Price: US $22.95/Dfl. 55.00

Vol. 2: LECTURES ON MICROECONOMIC THEORY

by E. MALINVAUD, *Institut National de la Statistique et des Etudes Economiques, Paris*
Translation by Mrs. A. SILVEY
1972 2nd repr. 1974 328 pages. Price: Clothbound: US $20.95/Dfl. 50.00
Paperback : US $14.75/Dfl. 35.00

Vol. 3: THE THEORY OF ECONOMIC PLANNING

by G. M. HEAL, *University of Sussex, England*
1973, 424 pages. Price: Clothbound: US $24.95/Dfl. 60.00
Paperback : US $16.75/Dfl. 40.00

Vol. 4: CAPITAL THEORY AND THE DISTRIBUTION OF INCOME

by C. J. BLISS, *University of Essex, Colchester, England*
1975, 392 pages. Price: Clothbound: US $24.95/Dfl. 60.00
Paperback : US $16.75/Dfl. 40.00

Vol. 5: APPLIED CONSUMPTION ANALYSIS

by LOUIS PHLIPS, *Université Catholique de Louvain, Belgique*
1974, 290 pages. Price: Clothbound: US $20.95/Dfl. 50.00
Paperback : US $14.75/Dfl. 35.00

Vol. 6: INTRODUCTION TO EQUILIBRIUM ANALYSIS

by W. HILDENBRAND, *University of Bonn, and* A. P. KIRMAN, *University of Warwick*
1975, 224 pages. Price: US $18.50/Dfl. 45.00

Vol. 7: FOUNDATIONS OF ECONOMETRICS

by ALBERT MADANSKY, *University of Chicago*
1975, about 300 pages, in preparation

North-Holland Publishing Company

P.O. Box 211, AMSTERDAM, The Netherlands.
52 Vanderbilt Avenue, NEW YORK NY 10017, U.S.A.

Prices are subject to change without prior notice.

0188

FROM THE UNIVERSITY OF MICHIGAN

New Books

Confronting Today's Major Economic Issues

Industrial Market Structure and Performance — 1960-1968

Daryl N. Winn

A rigorous statistical analysis of the relation between market structure and the actual performance of economic units. Prof. Winn focuses on the question: Is the consistent association of above-average returns with industrial market concentration necessarily a result of direct cause and effect (an assumption fundamental to current key recommendations for antitrust policy), or is it the result of multicollinearity among several phenomena? He explores the influences of capital intensity, growth, business risk, and distortions arising from the use of historical costs, in addition to examining the measures of profitability themselves for systematic bias in observed measures. An important contribution to the literature on public policy and antitrust economics.

Michigan Business Studies I, no. 2, new series; xi+ 236 pages; hardcover **$12.00**

The U.S. Medical Care Industry: The Economist's Point of View

Joseph Morreale, *editor*

Six expert economists assess issues such as: the main economic fallacies inherent in present policies on health care delivery in the United States; the impact of Blue Cross on the price of hospital care; the effectiveness of present subsidies of nurse training, and the essential criteria for an acceptable national health insurance plan. Invaluable background to the debate on National Health Insurance. Economists include Paul Feldstein, Harry Schwartz, Gerald Rosenthal, Donald E. Yett, Rueben A. Kessel, and Herbert Klarman.

Michigan Business Papers, no. 60; xvi + 108 pages; paper **$5.50**

Black Economic Development: Analysis and Implications

William L. Cash, Lucy R. Oliver, *editors*

A variety of perspectives from the literature of the seventies on the history and problems of black economic development in the United States as well as on actual and proposed programs for its improvement. Designed to provide the broadest possible spectrum of thinking about minority economics, this anthology juxtaposes divergent viewpoints from social philosophers, academicians, economists, and activists. A valuable textbook which includes discussion questions and suggestions for additional readings.

430 pages **$8.50**

Minorities at the Crossroads

Gerald F. Whittaker, *editor*

Selected lectures from the 1973 and 1974 Annual Symposiums on the State of the Black Economy. The twelve eminent speakers look at a variety of issues affecting the success of future minority economic growth — revenue sharing, the energy crisis, trade unions, education, income maintenance, and international trade.

xiv + 177 pages; paper **$2.50**

The Economic Effects of Multinational Corporations

Werner Sichel, *editor*

Do multinational corporations benefit or deter international economic units? Five prominent economists with divergent viewpoints confront this question. Includes an analysis of the effect of foreign multinationals in the U.S. economy.

Forthcoming

for orders and information write

Division of Research Publications Department
Graduate School of Business Administration
The University of Michigan
Ann Arbor, Michigan 48104

contemporary
ECONOMIC
problems and issues

HAILSTONES • MASTRIANNA • MARTIN

FOURTH EDITION

The new Fourth Edition of CONTEMPORARY ECONOMIC PROBLEMS AND ISSUES is a 15-chapter text that discusses just what the title says—current economic issues. This edition contains 2 new chapters on some controversial topics: "THE ENERGY CRISIS—REAL OR CONTRIVED?" and "FOOD—WHAT HAPPENED TO THE SURPLUSES?"

We know your students will find the topics interesting, but since seeing is believing, here is a chance to preview the Table of Contents:

Chapter
1 UNEMPLOYMENT—WHY AND FOR HOW LONG?
2 INFLATION—HOW MUCH AND WHERE DOES IT HURT?
3 WAGE AND PRICE CONTROLS—ARE THEY NECESSARY?
4 POVERTY—HAVE WE CURED IT?
5 INCOME—SHOULD IT BE GUARANTEED?
6 THE NATIONAL DEBT—WILL IT BANKRUPT US?
7 HEALTH INSURANCE—SHOULD UNCLE SAM PAY?
8 FOOD—WHAT HAPPENED TO THE SURPLUSES?
9 CLEAN AIR—ARE WE WILLING TO PAY THE PRICE?
10 URBAN RENEWAL—CAN IT SAVE OUR CITIES?
11 MASS TRANSIT—WILL IT GET US THERE?
12 GOING BROKE—IS REVENUE SHARING THE ANSWER?
13 GOLD—WHO NEEDS IT?
14 THE ENERGY CRISIS—REAL OR CONTRIVED?
15 THE POPULATION EXPLOSION—CAN THE WORLD FEED ITSELF

With topics like this, CONTEMPORARY ECONOMIC PROBLEMS AND ISSUES promises to capture students' interest and be a text they enjoy studying. Examination copies will be available in January. Write for yours today.

SOUTH-WESTERN PUBLISHING CO.

New from

COLUMBIA ESSAYS ON THE GREAT ECONOMISTS

DONALD DEWEY, General Editor

This series is designed to provide concise, authoritative studies of the men who have made the greatest contributions to modern economic thought. The following new titles are now available:

DAVID RICARDO

MICHAEL J. GOOTZEIT

Feeling that research efforts over the past fifty years have sufficiently covered Ricardo's writings on the production and distribution system, Professor Gootzeit, in this new essay, has chosen to emphasize the importance of the monetary system in Ricardo's economic thought.
$7.50, cloth; $1.95, paper

J. M. CLARK

C. ADDISON HICKMAN

John M. Clark's major contributions to modern economic theory—overhead costs, workable competition, and the acceleration principle—are accompanied here by a new look at his work in the field of welfare and his healing advice to economists. $7.50, cloth; $1.95, paper

JOHN MAYNARD KEYNES

HYMAN P. MINSKY

In this radical reevaluation of the work of John Maynard Keynes, an alternative interpretation of Keynes is presented which emphasizes the importance of uncertainty, speculation, and the complex financial structure in generating a "cyclical" rather than "an equilibrium growth" behavior of the economy. $12.00, cloth; $2.95, paper

THE TREATMENT OF MARKET POWER

Antitrust, Regulation, and Public Enterprise

WILLIAM G. SHEPHERD

In this innovative examination of the problems of power and equity in the American economy, Professor Shepherd first presents a new analysis of market power, reassessing its effects and permanence in the modern economy, and then offers suggestions as to how this power should be treated by public policies. $15.00

COLUMBIA UNIVERSITY PRESS

Address for orders: 136 South Broadway, Irvington, New York 10533

Columbia

NEW FROM THE NATIONAL BUREAU OF ECONOMIC RESEARCH

LONG WAVES IN URBAN DEVELOPMENT

MANUEL GOTTLIEB

Using detailed historical records and other available real estate statistics, the author examines long swings in urban development between the early 1840s and the late 1930s. This study brings out the essential characteristics of, and the interplay between, the different elements and processes at work in long swings.

Urban and Regional Studies, 4 $15.00

THE FORMATION AND STOCKS OF TOTAL CAPITAL

JOHN W. KENDRICK

Assisted by YVONNE LETHEM and JENNIFER ROWLEY

In this new book the author builds on his pioneering work in productivity to provide for the first time a comprehensive assembly of estimates of stock of all the types of capital, intangible as well as tangible, resulting from investments.

General Series, 100 $15.00

FOREIGN TRADE REGIMES AND ECONOMIC DEVELOPMENT

JAGDISH N. BHAGWATI and ANNE O. KRUEGER, Editors

This ambitious research project analyzes the phenomena of exchange control, liberalization, and economic development in less-developed countries. The three major topics covered in each country studied are: the anatomy of control, the liberalization episode, and growth relationships.

New Volumes in the series are:

Volume VI: India
by Jagdish N. Bhagwati and T. N. Srinivasin
$15.00, cloth; $5.00, paper

Volume VII: South Korea
by Charles R. Frank, Jr., Kwang Suk Kim, and Larry Westphal $15.00

A Special Conference Series on Foreign Trade Regimes and Economic Development

COLUMBIA UNIVERSITY PRESS

Address for orders: 136 South Broadway, Irvington, New York 10533

From The Free Press

THE DISADVANTAGED CONSUMER

Alan R. Andreasen, University of Illinois at Urbana

Do the disadvantaged—the poor, the racial minorities, the elderly, the uneducated, the non-English speaking—get their money's worth in the marketplace? Are they victimized by the operating practices of inner city businesses? Have they been helped or hindered by the consumer movement?

This is single best source of information on the special consumer problems of the disadvantaged, providing a systematic, documented analysis of the discrimination they face. Andreasen examines in detail the causes of their disadvantage in the marketplace, and proposes solutions to eliminate or moderate their consumer problems.

355 pages $14.95

MARKETS AND HIERARCHIES: ANALYSIS AND ANTITRUST IMPLICATIONS

A Study in the Economics of Internal Organization

Oliver E. Williamson, University of Pennsylvania

A ground-breaking study of economic organization and antitrust enforcement. Beginning with simple forms of internal firm organization, the author examines increasingly complex forms of economic hierarchies. Antitrust enforcement implications are spelled out with respect to vertical integration, conglomerate organization, organizational innovations, dominant firm industries, and oligopoly. "...a powerful addition to our knowledge ...a stimulus to further research."—Kenneth J. Arrow, Harvard University

288 pages $14.95

THE UNCERTAIN SEARCH FOR ENVIRONMENTAL QUALITY

A Case Study in the Failure of Modern Policy Making

Bruce A. Ackerman and Susan Rose Ackerman, both, Yale University; James W. Sawyer, Jr.; Dale W. Henderson

"...an extraordinary study of the policy-making process. Better than any other analysis that I know of, it deals with the complex interaction among technical, economic, bureaucratic, political, and legal factors that are characteristic of the policy process....A landmark analysis which will be widely used for years to come...."

—Henry S. Rowen, Professor of Public Management, Stanford University

367 pages $13.95

ECONOMIC MAN

The Anthropology of Economics

Harold K. Schneider, Indiana University

Drawing upon the concepts and theories of microeconomics, this book demonstrates their relevance for economic anthropology. It suggests a mode of analysis for understanding not only western and non-western economic systems but also social behavior and activity. "...challenges the bias against the concept of economic man as narrow and greedy."—*Journal of Economic Literature*

288 pages $8.95

THE FREE PRESS
A DIVISION OF MACMILLAN PUBLISHING CO., INC.
100D Brown Street, Riverside, New Jersey 08075

PUBLICATIONS FROM UNITED NATIONS

World Economic Survey 1974 Part I

Mid-term Review and Appraisal of Progress in the Implementation of the International Development Strategy.

Presents an over-all analysis of economic and social performance in the developing countries and of implementation of the Strategy by the developed market economies and the centrally planned economies.

Order No. E.75.II.C.1 $9.00

Ground-Water Storage and Artificial Recharge (Natural Resources Series No. 2)

Order No. E.74.II.A.11 $14.00

Management of International Water Resources: Institutional and Legal Aspects

Order No. E.75.II.A.2 $11.00

A New United Nations Structure for Global Economic Co-operation

Order No. E.75.II.A.7 $7.00

Profiles of Manufacturing Establishments Vol. IV: Three-Country Study on Export-Oriented Industries

Order No. E.74.II.B.13 $11.00

The Growth of World Industry 1973, Vol. II Commodity Production Data 1964-1973

Order No. E.75.XVII.4 Cloth $30.00

Summary of the Hearings Before the Group of Eminent Persons to Study the Impact of Multinational Corporations on Development and on International Relations

The hearings which constituted a novel approach for the United Nations, were described by the Group of Eminent Persons in their report as "a most useful source of information, as well as a valuable occasion to test ideas". The present document is published in the light of that opinion, in the belief that it will be of assistance to Governments and to the public in the further elucidation of this complicated issue.

Order No. E.74.II.A.9 $18.00

Industrial Development Survey. Special Issue for the Second General Conference of UNIDO

Order No. E.74.II.B.14 $11.00

Request our catalogues, which give a complete description of the United Nations publishing programme and are available free of charge:

- Check List of Books in Print 1975
- United Nations Official Records 1962-1972
  Supplement 1970-1972
- Standing Order Brochure
- United Nations Periodicals 1975

United Nations Publications, Room LX-2300, New York, N.Y. 10017
or
Palais des Nations, 1211 Geneva 10, Switzerland

REYNOLDS

for timeliness and quality of instruction in the first course in Economics

Available in February

TWO PAPERBACK VOLUMES

MICROECONOMICS: Analysis and Policy
MACROECONOMICS: Analysis and Policy
Revised Editions

by Lloyd G. Reynolds, *Yale University*

The emphasis of the new revised editions is on the *usefulness* of economic reasoning in analyzing policy issues—half of the material deals with "real world" problems.

Readable, teachable, and *learnable,* each volume is short and accurate allowing time for assignment flexibility.

There are no roadblocks, detours, or rough corners in the smooth logical presentations. Each chapter begins with introductory material designed to provide the student with significant guide lines to chapter content. A new "Think about This" section at the end of each chapter poses a relevant Key problem for student discussion.

Each volume has its own separate supporting materials constituting a compact, individual, mini-package. For Micro and Macro each, there is a Study Guide and Workbook by Richard Porter and Dennis Starleaf, an Instructor's Manual and a Test Bank.

Examination copies for adoption consideration available on request; please indicate course title and text presently used.

RICHARD D. IRWIN, INC. — Homewood, Illinois 60430